

Studying $Z/\gamma^* + \text{Jet}$ Production in Proton-Antiproton Collisions at $\sqrt{s} = 1.96 \text{ TeV}$

Dissertation zur Erlangung der Doktorwürde
Vorgelegt von

Henrik Wold Nilsen



Fakultät für Mathematik und Physik
Albert-Ludwigs-Universität Freiburg

Dekan:	Prof. Dr. Kay Königsmann
Leiter der Arbeit:	Prof. Dr. Karl Jakobs
Referent:	Prof. Dr. Karl Jakobs
Koreferent:	Prof. Dr. Ulrich Landgraf
Mündliche Prüfung:	30. Juli 2009

Contents

1	Introduction	1
2	The Standard Model of Particle Physics	3
2.1	Particle Content	3
2.2	Lagrangian	3
2.2.1	Local Gauge Invariance and Minimal Substitution	5
2.2.2	$SU(2)_L \times U(1)_Y$ Symmetry – Electroweak Unification	6
2.2.3	$SU(3)$ symmetry – QCD	7
2.2.4	Higgs Sector	7
2.3	From Lagrangian to Collision Rates	9
2.3.1	Calculating Cross Sections	10
2.3.2	Luminosity	10
2.4	Renormalization and Running Coupling Constants	11
2.5	Detecting the Higgs boson	14
2.6	Problems of the Standard Model	15
2.7	Extensions to the Standard Model	17
3	Phenomenology of $p\bar{p}$ Collisions	19
3.1	Fixed-Order Calculations	20
3.2	Including Terms to all Orders	20
3.2.1	Analytical Resummation	20
3.2.2	Parton Shower	21
3.3	Combining Fixed-Order and All-Order Calculations	23
3.3.1	Combining Fixed-Order and Analytical Resummation Calculations . .	23
3.3.2	Combining Fixed-Order Calculations and Parton-Shower Algorithms .	23
3.4	From Partons to Hadrons	26
3.4.1	Hadronization	26
3.4.2	Parton Distribution Functions	27
3.4.3	Intrinsic p_T	28
3.5	Underlying Event	28
3.5.1	Multiple Parton Interactions	28
3.5.2	Beam Remnants	29
3.5.3	Multiple $p\bar{p}$ Interactions	29
3.6	Event Generators	29
3.6.1	MCFM	30
3.6.2	PYTHIA	30
3.6.3	HERWIG and JIMMY	30
3.6.4	SHERPA	30
3.6.5	ALPGEN+PYTHIA and ALPGEN+HERWIG+JIMMY	30
3.6.6	MC@NLO	30

3.7	Connecting Partons, Particles and Energy-Deposits in the Detector	31
3.7.1	Cone Jet Algorithms	32
3.7.2	Other Jet Algorithms	33
4	Experimental Setup: Tevatron and the DØ Detector	37
4.1	The Tevatron Accelerator	37
4.2	The DØ Detector	38
4.2.1	Central Tracking System	39
4.2.2	Calorimeter System	42
4.2.3	Muons System	43
4.2.4	Luminosity Counter	43
4.2.5	Trigger and Data Acquisition System	44
4.2.6	Offline Reconstruction	45
4.3	Simulation of the Experimental Setup	45
5	Event Reconstruction and Object Identification	47
5.1	Tracks	47
5.2	Vertex Reconstruction	48
5.3	Calorimeter Zero Suppression	49
5.4	Electrons	49
5.4.1	Electron Identification	49
5.4.2	Electron Identification Efficiency	50
5.4.3	Electron Energy Scale and Resolution	53
5.5	Jets	53
5.5.1	Jet Identification	53
5.5.2	Jet Identification Efficiency	54
5.5.3	Jet Energy Scale from γ +jet production	55
5.5.4	Jet Energy Scale and Resolution from Z +jet production	56
6	Event Samples from Data and Simulation	61
6.1	Data Sample	61
6.1.1	Data Quality	61
6.1.2	Integrated Luminosity	61
6.2	Simulated Event Samples	61
7	Detector-Level Studies of Z/γ^*+jets production	63
7.1	Event Selection	63
7.2	Correcting the Performance of the Simulated Detector	64
7.3	Detector-level Comparisons: Data, PYTHIA and SHERPA	67
7.4	Conclusions	70
7.5	Limitations of Detector-level Comparisons	71
8	Correcting Data to Particle Level	79
8.1	Ansatz-Function-Based Method	80
8.2	Event-Generator-Based Ansatz Method	80
8.3	Full-Detector-Simulation Method	81
8.4	Regularized versus Bin-By-Bin Unfolding	81

9	Measurements of Differential Cross Sections in Z/γ^*+jets Events	87
9.1	Event Selection	87
9.2	Normalizing the Observables	88
9.3	Backgrounds	88
9.4	Correction Data to Particle Level	92
9.5	Detector-Level versus Particle-Level Electrons	93
9.6	Consistency checks of the Jet Definition	94
9.7	Choice of binning	95
9.8	Distinguishing Detector and Generator Effects	97
9.9	Collider-specific corrections	99
9.10	Detector-specific Corrections related to Electrons	100
9.11	Measuring the Inclusive $Z/\gamma^* \rightarrow e^+e^-$ Cross Section	101
9.12	Detector-specific Corrections related to Jets	102
9.12.1	Jet Properties	103
9.13	Jet Dependency of Electron Corrections	104
9.13.1	Jet Dependency of the Trigger Efficiency	108
9.13.2	Jet Dependency of the Electron ID Efficiency	110
9.14	Correcting Kinematic Properties of the Simulated Event Sample	112
9.15	Jet Energy Resolution Changes the p_T^{jet} Hierarchy	122
9.16	Total correction factors	124
9.17	Extrapolation for the $\Delta R(e, \text{jet}) > 0.4$ cut	124
9.18	Extrapolation for the lepton cuts	126
9.19	Uncertainties	127
9.20	Final Results	127
9.21	Comparisons with existing measurements	132
9.22	Measurement of the Differential $\Delta\phi(\text{di-electron, jet})$ Distribution	135
10	Comparing Theory Predictions with the Measurements	145
10.1	Theory Predictions	145
10.1.1	SHERPA	145
10.1.2	ALPGEN+PYTHIA and ALPGEN+HERWIG	146
10.1.3	PYTHIA	146
10.1.4	HERWIG+JIMMY	146
10.1.5	MC@NLO+JIMMY	146
10.1.6	MCFM	146
10.2	Comparing the Predictions with the p_T^{jet} Measurements	147
10.2.1	Fixed-order Calculations at LO and NLO from MCFM	148
10.2.2	SHERPA and ALPGEN+PYTHIA	148
10.2.3	MC@NLO	150
10.2.4	PYTHIA and HERWIG	150
10.2.5	The Relationship between MC@NLO, SHERPA, PYTHIA and HERWIG	152
10.2.6	Adjusting the parameters of ALPGEN+PYTHIA and SHERPA to describe the p_T^{jet} Measurements	155
10.3	Comparing the Predictions with the $\Delta\phi(\text{di-electron, jet})$ Measurements	156
10.4	The Sensitivity of $\Delta\phi(\text{jet, jet})$ to Jets from Multiple Parton Interactions	158
11	From the Tevatron to the LHC	161

11.1 p_T^{jet} spectra	161
11.2 Using $\Delta\phi(\text{dielectron, jet})$ to constrain the Cross Section for Multiple Parton Interactions	164
12 Summary and Conclusions	167
Lists of Tables and Figures	I
Bibliography	XIV
Acknowledgements	XV

Chapter 1

Introduction

The production of jets in association with a Z/γ^* boson is an example of an important class of processes at hadron colliders, namely vector boson+jet (V +jet) production. Comparisons of measurements of this class of processes with theory predictions constitute an important, fundamental test of the Standard Model of particle physics, and of the theory of QCD in particular. While having a smaller cross section than other V +jet processes, $Z/\gamma^*(\rightarrow e^+e^-)$ +jets production, with $Z/\gamma^* \rightarrow e^+e^-/\mu^+\mu^-$, has a distinct experimental signature allowing for measurements characterized by low backgrounds and a direct, precise measurement of the properties of the decay products of the Z/γ^* boson.

In this thesis, several new measurements of the properties of jets produced in association with a Z/γ^* boson in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV are presented. The cross section for $Z/\gamma^*(\rightarrow e^+e^-)+N$ jet production ($N \leq 3$) is measured, differential in the transverse momentum of the N^{th} jet in the event, normalized to the inclusive Z/γ^* cross section. Also, the cross section for $Z/\gamma^*(\rightarrow e^+e^-)+N$ jets ($N \geq 1$) is measured, differential in the difference in azimuthal angle between the di-electron system and any jet in the event, normalized to unity. The data used in the measurements were collected by the DØ experiment located at the Tevatron Collider of the Fermi National Accelerator Laboratory and correspond to an integrated luminosity of 1.04 fb^{-1} .

The measured jet transverse momentum spectra are compared with the predictions of perturbative calculations at the next-to-leading order in the strong coupling constant. Given the low sensitivity of the calculations to model parameters, these comparisons represent a stringent test of perturbative QCD.

One of the main goals currently being pursued in particle physics is the discovery of the only particle predicted by the Standard Model which has so far no been detected experimentally, namely the Higgs boson. It is assumed that the ATLAS and CMS experiments located at the Large Hadron Collider (LHC), a proton-proton collider at $\sqrt{s} = 14$ TeV, will be able to detect the Higgs boson, or rule out its existence, within the next few years. The collisions delivered by the LHC will also be used to perform a long range of searches for other new particles, for instance particles predicted by models based on the principle of supersymmetry. The associated production of vector bosons with jets has relatively large production rates at the LHC and can produce a long list of different final states which can include charged leptons, missing transverse energy, as well as light- and heavy-flavour jets. This makes V +jet production a major source of background events to many searches for new particles.

Most techniques used for estimating the expected number of background events to searches rely on passing the stable final-state particles of simulated hadron collisions generated using a so-called *event generator* code, through a simulation of the experimental detector system. The development of event generators which are capable of reliably predicting the properties of jets produced in association with a core process, e.g. the production of a vector boson, has been the subject of a large amount of research activity during the last ten years. These efforts

have led to the appearance of the CKKW and MLM algorithms which are implemented in several event generators, among them SHERPA and ALPGEN+PYTHIA. The large data sample collected by the DØ experiment during Run II offers an excellent opportunity for validating these new event generators against experimental measurements of V +jet production. As argued above, the $Z/\gamma^*(\rightarrow e^+e^-)$ +jets process offers the combination of a clean experimental signature and large production rates, making it the process of choice for these studies.

In Chapter 2 of this thesis, a short summary of the Standard Model of particle physics is given. The various models used for deriving observable predictions from the Standard Model are reviewed in Chapter 3, with particular emphasis on the description of jets in V +jet production. Chapter 4 gives an overview of the Tevatron $p\bar{p}$ collider and the DØ experiment. The process of reconstructing physics objects and their energies from the signals registered in the DØ detector is summarized in Chapter 5, with particular emphasis on electrons and jets. A description of the analyzed data sets is given in Chapter 6. The predictions of the SHERPA and PYTHIA event generators of jet-related observables for $Z/\gamma^*(\rightarrow e^+e^-)$ +jet production are compared with distributions reconstructed in DØ data in Chapter 7. To facilitate direct comparisons between the DØ data and theory predictions, the distributions reconstructed in data must be corrected for the impact of the detector, and various techniques for performing these corrections are discussed in Chapter 8. One of these techniques is applied in Chapter 9 to measure the transverse-momentum spectra of the three leading jets in $Z/\gamma^*(\rightarrow e^+e^-)$ +jet production, as well as the shape of the spectrum of differences in azimuthal angle between the Z/γ^* and any jet in the event. In Chapter 10 these measurements are compared with the predictions of various theory calculations and commonly used event-generator models. Detailed studies of the relative performance of the different event-generator models are presented. Finally, in Chapter 11, similarities and differences between Z/γ^* +jet production at the Tevatron and the LHC are discussed.

Chapter 2

The Standard Model of Particle Physics

The Standard Model of particle physics is a theory which successfully describes the results of all particle physics experiments to date¹. It was formulated in the 1960s and 1970s, and has predicted the existence of several new particles which have later been found in experiments, e.g. the top quark and the W and Z bosons.

The present Chapter contains a brief summary of the Standard Model of particle physics, and a motivation for the measurements which constitute the main part of this thesis.

2.1 Particle Content

The particles currently thought to be fundamental can be divided into two categories: matter particles and force carriers. The matter particles can be thought of as the building blocks of matter. The force particles mediate interactions between matter particles. The Standard Model is a quantum field theory based on a $SU(3) \otimes SU(2) \otimes U(1)$ gauge symmetry. Each particle is fully specified by quantum numbers and its mass:

- the electric (Q) and weak (I_3) charges specify interaction strength of the particle with the force particles of the electroweak force (the photon, the Z and W^\pm);
- similarly, the color charge defines the interaction strength of the particle with the force particles of the strong force (gluons);
- the mass of a particle is given by its interaction with the Higgs field.

For instance, an electron is a spin- $\frac{1}{2}$ particle which has $Q = -1$, $I_3 = -\frac{1}{2}$, no color charge and a mass of 511 keV.

Particles with integer (half-integer) spin are referred to as *bosons* (*fermions*). All Standard Model force-mediating particles are spin-1 bosons, whereas the matter particles are spin- $\frac{1}{2}$ fermions. Fermions can be divided into particles with zero color charge, called leptons, and particles with non-zero color charge, called quarks. A summary of the various fermions, their masses and their quantum numbers is given in Tabs. 2.1 and 2.2.

2.2 Lagrangian

The Standard Model is a quantum field theory based on the principle of gauge invariance. It is specified in terms of a Lagrangian density, \mathcal{L}_{SM} , which determines the equations of motion through a minimization of the action, $S = \int d^4x \mathcal{L}_{\text{SM}}$. The Lagrangian is required to be

¹Apart from the existence of massive neutrinos.

Leptons			Quarks		
Name	Symbol	Mass	Name	Symbol	Mass
electron neutrino	ν_e	$< 3 \text{ eV}$	up	u	1.5 to 3.3 MeV
muon neutrino	ν_μ	$< 0.2 \text{ MeV}$	down	d	2.5 to 5.0 MeV
tau neutrino	ν_τ	$< 18.2 \text{ MeV}$	strange	s	70 to 130 MeV
electron	e	511.0 keV	charm	c	1.1 to 1.3 GeV
muon	μ	105.7 MeV	bottom	b	4.1 to 4.4 GeV
tau	τ	1.777 GeV	top	t	$171.2 \pm 2.1 \text{ GeV}$

Table 2.1: The Standard Model fermions and their masses (from Ref. [1]). Since quarks are not observable as free particles their exact masses are model dependent. Consequently, a mass interval is specified, corresponding to a range of different models. Throughout this thesis natural units are used. In this system, the reduced Planck constant \hbar and the speed of light c are $\hbar \equiv c \equiv 1$ and energies are given in units of electron-volt (eV).

Fermions	Generation			Q [e]	I_3	$SU(3)$	$SU(2)$	$U(1)$
Quarks	$\begin{pmatrix} u \\ d' \end{pmatrix}_L$	$\begin{pmatrix} c \\ s' \end{pmatrix}_L$	$\begin{pmatrix} t \\ b' \end{pmatrix}_L$	$\frac{2}{3}$	$\frac{1}{2}$	3	2	$\frac{1}{3}$
	u_R	c_R	t_R	$-\frac{1}{3}$	$-\frac{1}{2}$	3	1	$\frac{4}{3}$
	d_R	s_R	b_R	$\frac{2}{3}$	0	3	1	$-\frac{2}{3}$
Leptons	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$	$-\frac{1}{3}$	0	3	1	$-\frac{2}{3}$
	e_R	μ_R	τ_R	0	$\frac{1}{2}$	1	2	-1
				-1	$-\frac{1}{2}$	1	1	-2

Table 2.2: An overview of the fermions of the Standard Model and their quantum numbers. The electroweak force distinguishes between so-called left-handed (L) and right-handed (R) eigenstates. The quark eigenstates of the electroweak interaction, denoted d' , s' , b' , are related to the mass eigenstates d , s and b via the Cabibbo–Kobayashi–Maskawa–Matrix. Given that neutrino oscillations have been experimentally confirmed, neutrinos are known to have non-zero mass [2]. Consequently, they also have right-handed contributions.

invariant under Lorentz transformations, so as to give a relativistic theory, and it contains terms describing free fields as well as interactions between different fields.

The Standard Model Lagrangian can be written as

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}. \quad (2.1)$$

The two first terms, \mathcal{L}_{EW} and \mathcal{L}_{QCD} , describe free fermions, free bosons (*gauge bosons*) associated with the $SU(2) \times U(1)$ and $SU(3)$ gauge symmetries, the interaction between fermions and gauge bosons, and the interactions among gauge bosons themselves. The terms $\mathcal{L}_{\text{Higgs}}$

and $\mathcal{L}_{\text{Yukawa}}$ introduce the Higgs particle and non-zero gauge-boson and fermion masses. In the next sections, a short review of the principle of local gauge invariance and minimal substitution is given, followed by a discussion of each term in Eq. (2.1).

2.2.1 Local Gauge Invariance and Minimal Substitution

In the absence of interactions, massless fermions can be described by the Lagrangian

$$\mathcal{L} = \bar{\Psi}(i\gamma^\mu \partial_\mu)\Psi, \quad (2.2)$$

with γ^μ being Dirac matrices, Ψ a fermionic field and $\bar{\Psi}$ its adjoint field. The corresponding equation of motion is the Dirac equation which describes relativistic spin- $\frac{1}{2}$ fermions. The guiding principles for introducing additional terms in the Lagrangian to describe interactions are those of symmetry groups and gauge invariance. These principles were known to give the correct interaction terms for electromagnetism and have been successfully applied in understanding the electroweak force and the strong force. The Standard Model Lagrangian is required to be invariant under local phase or gauge transformations arising from the $U(1)$, $SU(2)$ and $SU(3)$ symmetry groups:

$$\begin{aligned} \Psi &\rightarrow \Psi' = e^{-i\tau_j \omega_j(x)} \Psi \\ \bar{\Psi} &\rightarrow \bar{\Psi}' = e^{i\tau_j \omega_j(x)} \bar{\Psi} \end{aligned} \quad (2.3)$$

where τ_j is the j^{th} generator of the group, and $\omega_j(x)$ is an arbitrary real differential function. For $U(1)$ there is one generator, a scalar constant, whereas for $SU(N)$, $N = 2, 3$, there are $N^2 - 1$ generators which can be represented as traceless $N \times N$ matrices. The requirement that the Lagrangian (Eq. (2.2)) has a local gauge invariance, i.e. allowing ω_j in Eq. (2.3) to be a function of space-time, leads to the introduction of new field(s), A_i^μ , through a so-called *minimal substitution*:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu - ig \frac{\tau^i}{2} A_{i\mu}, \quad (2.4)$$

where τ^i are the generator(s) of the symmetry group in question, g is a coupling constant, and A_i^μ are spin-1 gauge boson fields. The gauge transformation properties of the gauge fields are defined to cancel the term arising from the partial derivative in Eq. (2.2) acting on $\omega_j(x)$, defined in Eq. (2.3). The substitution defined by Eq. (2.4) gives one new vector field, with an associated vector boson, per generator of the symmetry group. The vector field(s) per construction interact with the fermionic field of Eq. (2.2), and can be interpreted as the messenger particle of the force arising from the gauge symmetry. For $SU(2) \times U(1)$ the gauge fields describe the electroweak force, whereas the $SU(3)$ gauge fields describe the strong force.

Gauge fields enter the Lagrangian through Eqs. (2.2) and (2.4), but also through kinematic terms describing the gauge fields in absence of interactions with fermions, of the form

$$-\frac{1}{4} F_{i\mu\nu} F_i^{\mu\nu}. \quad (2.5)$$

Here $F_{i\mu\nu}$ denotes the field tensor of the gauge field $A_{i\mu}(x)$, and it is defined as

$$F_{i\mu\nu} = (\partial_\mu A_{i\nu}(x) - \partial_\nu A_{i\mu}(x)) + gf_{ijk} A_{j\mu} A_{k\nu}, \quad (2.6)$$

where g is the coupling constant introduced in Eq. (2.4) and f_{ijk} is the structure constants which specify the group. The structure constants define the commutation relations of the generators of the symmetry group through

$$[\tau_i, \tau_j] = if_{ijk}\tau_k. \quad (2.7)$$

2.2.2 $SU(2)_L \times U(1)_Y$ Symmetry – Electroweak Unification

The $SU(2)_L \times U(1)_Y$ symmetry of the Standard Model gives rise to the unification of the electromagnetic and weak forces through the Glashow-Weinberg-Salam model. The $SU(2)_L$ symmetry distinguishes between *left-handed* and *right-handed* fermion states, defined as

$$\begin{aligned} \Psi_L &= P_L \Psi \\ \Psi_R &= P_R \Psi, \end{aligned} \quad (2.8)$$

where P_L and P_R are projection operators which satisfy $P_L \cdot P_R = P_R \cdot P_L = 0$ and $P_L + P_R = 1$, implying $P_L = P_L^2$, $P_R = P_R^2$. Left-handed fermions transform as a doublet under $SU(2)_L$, i.e. have weak isospin $I_3 = \pm \frac{1}{2}$, whereas right-handed fermions are singlets, i.e. have² $I_3 = 0$. The gauge bosons of the $SU(2)_L$ group couple only to left-handed fermions, and the Lagrangian has the form

$$\mathcal{L}_{\text{EW}} = i\bar{L}\gamma^\mu D_\mu L + i\bar{R}\gamma^\mu D_\mu R - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W_{\mu\nu}^i W_i^{\mu\nu}, \quad (2.9)$$

where L denotes left-handed fermion doublets and R the corresponding right-handed singlets (see Tab. 2.2). The $SU(2)_L \times U(1)_Y$ covariant derivative is given in analogy with Eq. (2.4) as

$$D_\mu = \partial_\mu + ig_1 \frac{Y}{2} B_\mu + ig_2 \tau_i W_\mu^i, \quad (2.10)$$

with B_μ being the $U(1)_Y$ gauge field, g_1 the $U(1)_Y$ coupling constant and Y the $U(1)_Y$ group generator. Similarly, for $SU(2)_L$, W_μ^i ($i = 1, 2, 3$) denotes the gauge fields, g_2 the coupling constant, and τ_i the three generators of the symmetry group. The boson field tensors, $B_{\mu\nu}$ and $W_{\mu\nu}^i$, are defined by Eq. (2.6). The $U(1)_Y$ structure constants are zero, corresponding to a commuting gauge field. The electroweak Lagrangian, \mathcal{L}_{EW} , contains terms which are at most bi-linear in B_μ , meaning that the $U(1)_Y$ gauge field couples to fermions, but not to itself. For $SU(2)_L$, the structure constants are given as the completely anti-symmetric tensor, ϵ_{ijk} . The non-zero structure constants yield non-commutative gauge (Eq. (2.7)) fields and tri-linear and quartic gauge terms (Eqs. (2.5) and (2.6)) in \mathcal{L}_{EW} , corresponding to gauge bosons which couple to themselves. Two of the $SU(2)_L$ gauge fields can be combined linearly into the physical, charged fields

$$W_\mu^\pm = \frac{1}{\sqrt{2}}[W_\mu^1 \mp iW_\mu^2] \quad (2.11)$$

whose quanta are observed as W^\pm bosons. The remaining neutral $SU(2)_L$ field mixes with the $U(1)_Y$ field through the weak mixing angle, θ_W , giving the fields A_μ and Z_μ , defined as

$$\begin{aligned} A_\mu &= B_\mu \cos \theta_W + W_\mu^3 \sin \theta_W, \\ Z_\mu &= -B_\mu \sin \theta_W + W_\mu^3 \cos \theta_W. \end{aligned} \quad (2.12)$$

²The analogy for electromagnetism would be that left-handed fermions have a non-zero electric charge, whereas right-handed fermions are charge neutral.

The weak mixing angle is defined by requiring the quanta of the A_μ field to correspond to photons. This is fulfilled by the choice $g_1 \sin \theta_W = g_2 \cos \theta_W = e$, with e the absolute value of the electric charge of an electron. The quanta of the Z_μ field are observed as the Z boson. The charge corresponding to the $U(1)_Y$ symmetry is called hyper-charge, Y , whereas for $SU(2)_L$ it is referred to as weak isospin, I_3 . These charges are related to the electrical charge, Q , through

$$Q = \frac{1}{2}Y + I_3.$$

2.2.3 $SU(3)$ symmetry – QCD

The $SU(3)$ symmetry of the Standard Model describes the strong force between quarks. The theory of $SU(3)$ is referred to as QCD, and its Lagrangian is given as

$$\mathcal{L}_{\text{QCD}} = i\bar{q}_f \gamma^\mu D_\mu q_f - \frac{1}{4} G_{\mu\nu}^i G_i^{\mu\nu}, \quad (2.13)$$

with q_f being a color triplet of quarks of flavor f . The covariant derivative, D_μ is given by

$$D_\mu = \partial_\mu - ig_3 \frac{\lambda_i}{2} G_\mu^i. \quad (2.14)$$

Here, λ_i ($i = 1 \dots 8$) are the eight $SU(3)$ generators and G_μ^i the corresponding gauge fields, with g_3 being the coupling constant. The quanta of the $SU(3)$ fields are called gluons.

The $SU(3)$ generators commute according to Eq. (2.6) with the structure constants given by the completely anti-symmetric f_{ijk} , with the non-zero components being defined through

$$\begin{aligned} f_{123} &= 1, \\ f_{147} = f_{246} = f_{257} = f_{345} = f_{516} = f_{637} &= \frac{1}{2}, \\ f_{458} = f_{678} &= \frac{\sqrt{3}}{2}. \end{aligned} \quad (2.15)$$

As for $SU(2)_L$, the non-zero structure constants of $SU(3)$ give rise to tri-linear and quartic terms in \mathcal{L}_{QCD} . This corresponds to gluons carrying a non-zero color charge. Since the $SU(3)$ color symmetry is unbroken gluons have zero mass.

2.2.4 Higgs Sector

Introducing mass terms as $m^2 \bar{\Psi} \Psi$, where Ψ denotes any Standard Model field, would break the $SU(2)_L$ invariance of the Standard Model. This follows since $\bar{\Psi} \Psi = \bar{\Psi} (P_L + P_R) \Psi = \bar{\Psi} (P_L^2 + P_R^2) \Psi = \bar{\Psi}_R \Psi_L + \bar{\Psi}_L \Psi_R$, and since left-handed and right-handed fields transform differently under the $SU(2)_L$ gauge transformation defined by Eq. (2.3):

$$\begin{aligned} \Psi_L &\rightarrow \Psi'_L = e^{i\tau_j \omega_j(x)} \Psi_L \\ \Psi_R &\rightarrow \Psi'_R = \Psi_R. \end{aligned} \quad (2.16)$$

Consequently, the term $m^2 \bar{\Psi} \Psi$ is not gauge invariant and cannot be included in the Lagrangian. Instead, mass terms are accommodated into the Standard Model by postulating the existence of a scalar bosonic field with a non-zero vacuum expectation value, referred to as the Higgs field. Unlike for the force-mediating fields introduced above, there is no

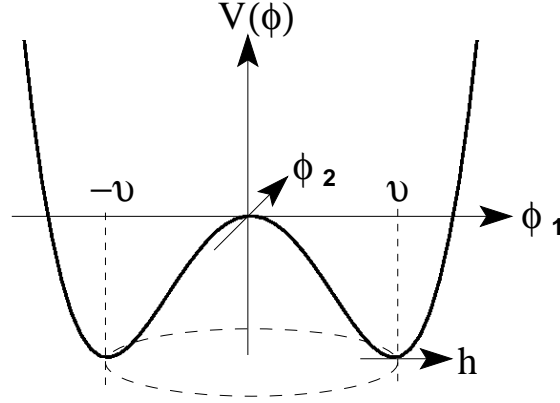


Figure 2.1: The Standard Model Higgs potential for a complex field Φ .

gauge mechanism responsible for the Higgs field. The Higgs field is a scalar, complex $SU(2)_L$ doublet

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad (2.17)$$

described by the Lagrangian

$$\mathcal{L}_{\text{Higgs}} = |D_\mu \Phi|^2 - V(\Phi), \quad V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2, \quad (2.18)$$

with $\lambda > 0$ for the scalar potential $V(\Phi)$ to be bounded from below. The covariant derivative of $SU(2)_L \times U(1)_Y$, D_μ , defined in Eq. (2.10), introduces coupling terms between the Higgs field and the gauge bosons. For $\mu^2 < 0$, the ground state, Φ_0 , can be written as

$$\Phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad v = \sqrt{\frac{-\mu^2}{\lambda}}, \quad (2.19)$$

corresponding to a non-vanishing vacuum expectation value

$$\langle 0 | \Phi^2 | 0 \rangle = \frac{v^2}{2}. \quad (2.20)$$

Whereas the Lagrangian is $SU(2)_L \times U(1)_Y$ symmetric, the ground state is not – this is referred to as *spontaneous symmetry breaking*. The field Φ can be expanded around the ground state as

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \eta_1 + i\eta_2 \\ v + h + i\eta_3 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix}. \quad (2.21)$$

The last equality follows from transforming from an arbitrary to the unitary gauge. The quanta of the h field are called Higgs bosons. In the unitary gauge the three η_i fields appear as longitudinal degrees of freedom for the Z and W^\pm gauge bosons. The three gauge bosons thus acquire masses, given as

$$m_W = \frac{1}{2} v g_2, \quad m_Z = \frac{m_W}{\cos \theta_W}, \quad (2.22)$$

In contrast, the quanta of the A_μ field, defined in Eq. (2.12), remain massless. The masses m_W and m_Z can be expressed in terms of well-measured experimental quantities, i.e. the gauge boson masses are predictions of the Standard Model, and these predictions are found to be in good agreement with current experimental data.

The Higgs boson also acquires a mass through its self-coupling terms, given as

$$m_H = \sqrt{-2\mu^2}. \quad (2.23)$$

Unlike m_W and m_Z , m_H is not predicted by the Standard Model since μ is essentially a free parameter, only required to satisfy $\mu^2 < 0$.

The masses of the Standard Model fermions can be described by introducing terms in the Lagrangian which couple the fermion fields to the Higgs field (so-called *Yukawa* terms). For example, the Yukawa terms for electrons are

$$g_e (\bar{L}\Phi e_R + \Phi^\dagger \bar{e}_R L) \quad (2.24)$$

where L denotes the first-generation left-handed lepton doublet and e_R the right-handed electron siglet (see Tab. 2.2). In the unitarity gauge this can be written as

$$\frac{g_e v}{\sqrt{2}} \bar{e}e + \frac{g_e}{\sqrt{2}} \bar{e}e h. \quad (2.25)$$

The first term is the electron mass term, and the second term specifies the coupling of the electron to the Higgs boson. The electron mass is

$$m_e = g_e v / \sqrt{2}, \quad (2.26)$$

with the electron coupling to the Higgs boson being

$$m_e / v. \quad (2.27)$$

Similar Yukawa terms are introduced for the other charged leptons and for the quarks. Analogous to Eq. (2.27), the couplings of the Higgs boson to fermions are always proportional to the fermion mass.

2.3 From Lagrangian to Collision Rates

The Standard Model not only predicts which particles exist, but also how often each particle type will be produced in a given experimental collider setup, e.g. the Tevatron or the LHC.

By convention, the description of production rates is based on a purely classical analogy: throw point-like objects (*probes*) towards a collection of n_1 non-overlapping circles (*targets*), each with an area σ , which are contained inside a larger circle (*pad*), with area A_{eff} . Each probe is guaranteed to hit inside the pad, and the probability that it hits a target is given by

$$P_{\text{hit}} = \frac{n_1 \cdot \sigma}{A_{\text{eff}}}. \quad (2.28)$$

Assuming that n_2 independent probes are thrown towards the pad with a frequency, f , the expected hit rate per second, \dot{N} , is given by

$$\dot{N} = f \cdot \frac{n_1 \cdot n_2}{A_{\text{eff}}} \cdot \sigma. \quad (2.29)$$

This picture is also used for collider experiments where bunches of particles with an effective area of A_{eff} containing n_1 and n_2 particles intersect with a frequency of f . The production rate of a process, e.g. $p\bar{p} \rightarrow Z/\gamma^* \rightarrow ee + X$, is given by the *cross section* of the process, σ , corresponding to the area of each target in the classical picture above.

2.3.1 Calculating Cross Sections

The cross section for a transition from two incoming particles into N outgoing particles can be expressed as

$$\sigma_{2 \rightarrow N} = \int |\mathcal{M}|^2 \times \mathcal{P}_{\text{kinematics}} \quad (2.30)$$

where $|\mathcal{M}|^2$ is the *squared matrix element* and $\mathcal{P}_{\text{kinematics}}$ is a *phase-space* factor. The squared matrix element can be determined from the Lagrangian using perturbation theory and the phase-space factor is given by the kinematic properties of the initial-state and final-state particles. The matrix element, \mathcal{M} , is normally written in terms of a perturbation series expansion in the coupling constant of the interaction. Following Feynman, each term in the series can be determined from graphical interaction diagrams, *Feynman diagrams*, using translation rules, *Feynman rules*, which are given by the Lagrangian. Including only terms to the lowest order in the coupling constant gives a *leading order* (LO) computation. Including one order more in the coupling constant gives a *next-to-leading order* (NLO) computation, and so on.

Cross sections, which have the dimensions of area, are normally quoted in units of *barn* (b), defined by

$$1 \text{ b} \equiv 10^{-24} \text{ cm}^2. \quad (2.31)$$

As an example, the cross section for $p\bar{p} \rightarrow Z/\gamma^* \rightarrow ee + X$ at the Tevatron collider (described in Chapt. 4), for $60 < M_{Z/\gamma^*} < 130$ GeV, is ≈ 250 pico-barn (pb). To appreciate how unlikely it is that this process takes place, it is useful to consider the classical analogy of a target area: assuming that the radius of a proton is 10^{-15} m, the “target area” of $p\bar{p} \rightarrow Z/\gamma^* \rightarrow ee + X$ interaction is $\approx 10^{12}$ times smaller than the area of a proton. Or, if one would collide a single $p\bar{p}$ pair head on once per second³, the average interaction rate would be one per 30k years.

2.3.2 Luminosity

The factorization of the production rate into a theory-dependent interaction part and a beam-parameter dependent part is formalised through

$$R = L \cdot \sigma, \quad (2.32)$$

with the *instantaneous luminosity* of the collider, L , being defined as

$$L = f \cdot \frac{n_1 \cdot n_2}{A_{\text{eff}}}. \quad (2.33)$$

Instantaneous luminosity is measured in units of $[\text{length}]^{-2} \cdot [\text{time}]^{-1}$, normally in units of $\text{cm}^{-2} \text{ s}^{-1}$. The effective area of each bunch is expressed in terms of the horizontal (σ_x) and the vertical (σ_y) bunch width under the assumption of a perfectly Gaussian bunch profile.

Whereas the instantaneous luminosity is a parameter of the experimental setup, the size of a data set is characterised by the integrated luminosity,

$$\mathcal{L} = \int L dt, \quad (2.34)$$

³Assuming $n_1 = n_2 = 1$, $A_{\text{eff}} = A_{\text{proton}} = 10^{-30} \text{ m}^2$ and $f = 1 \text{ Hz}$ in Eq. (2.29).

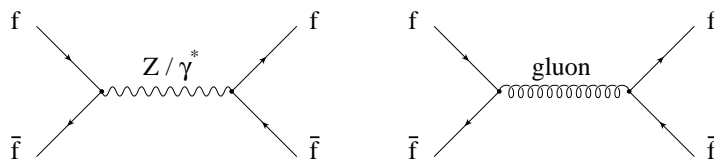


Figure 2.2: Leading-order diagrams for fermion+fermion \rightarrow fermion+fermion mediated by a Z/γ^* propagator (left) and a gluon propagator (right).

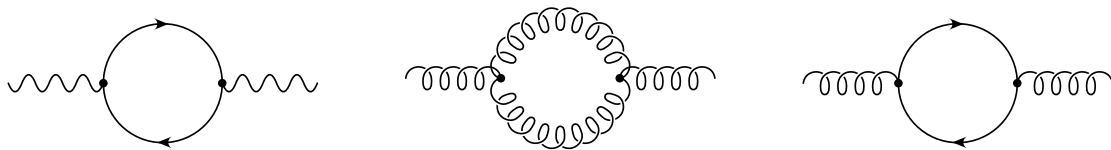


Figure 2.3: Fermion-loop diagrams contributing to the Z/γ^* propagator (left). Both fermion- and boson-loop diagrams contribute to the gluon propagator (center and right).

where the integral is performed over the time period when the data set was registered (*data-taking period*). A process with cross section σ is expected to have happened

$$N = \mathcal{L} \cdot \sigma \quad (2.35)$$

times during the data-taking period.

2.4 Renormalization and Running Coupling Constants

The coupling constants appearing in the Standard Model Lagrangian are dimensionless, real, positive numbers. When computing cross sections, the inclusion of terms beyond leading order in the coupling constant leads to divergences which can be treated using the technique of *renormalization*. To illustrate such divergences, we consider the transition from two incoming fermions into two outgoing fermions via the interaction of a gauge boson. Leading order interaction terms correspond to Feynman diagrams with the exchange of a single gauge boson (see Fig. 2.2). Higher-order terms receive contributions from Feynman diagrams with additional outgoing particles (*real* corrections), or with so-called *loops* (*virtual* corrections). A loop can contain all particle-antiparticle pairs which carry the charge to which the gauge boson couples. Examples of loop diagrams are given in Fig. 2.3 for the case of a Z/γ^* and a gluon mediated interaction. Only the momentum sum of the two loop particles is physically observable, not the momenta of each loop particle itself, and as a result the transition probability computation includes an integral over all possible loop particle momenta, and these integrals lead to a divergent result. Through renormalization, such divergences are avoided defining physical quantities, e.g. masses or coupling constants, as the combination of the *bare* quantity appearing in \mathcal{L}_{SM} and the divergent loop contribution. For *renormalizable* theories, like the Standard Model, the unobservable value of the bare quantity can be chosen such that loop divergences to all orders in perturbation theory are cancelled. As a consequence of renormalization, the value of the coupling constants depend on the energy-scale

(the *renormalization scale*, μ_R) which defines which higher-order contributions are absorbed into the redefinition of the quantities of the Lagrangian, and which are taken into account using perturbation theory. The renormalized coupling constants, referred to as the *running* coupling constants, obey the so-called *renormalization group equation*

$$Q^2 \frac{\partial \alpha}{\partial Q^2} = \beta(\alpha), \quad (2.36)$$

where the β -function can be expressed as a power-series in α using perturbation theory. Including only LO terms in α , i.e. assuming

$$\beta(\alpha) = -\beta_0 \alpha^2, \quad (2.37)$$

and solving Eq. (2.36) for α gives

$$\alpha(Q^2) = \frac{\alpha(\mu^2)}{1 + \beta_0 \alpha(\mu^2) \log \frac{Q^2}{\mu^2}}, \quad (2.38)$$

where the constant β_0 depends on the particles which can appear in the loops. For two electrically charged low-energetic particles interacting via a photon exchange, the regime of quantum electrodynamics, all electrically charged fermions can appear in the loops, yielding

$$\beta_0^{\text{QED}} = -\frac{1}{3\pi} < 0, \quad (2.39)$$

meaning that the coupling constant of QED increases with increasing momentum transfer. For two quarks interacting via gluon exchange the situation is different. As noted in Sect. 2.2.3, the non-zero structure constants of the $SU(3)$ group give non-commuting gauge fields and tri-linear and quartic gluon self-coupling terms. As a consequence, the loops along the gluon line can contain both quarks and gluons, giving

$$\beta_0^{\text{QCD}} = \frac{11 \cdot C_A - 2n_f}{12\pi}, \quad (2.40)$$

where the term $11 \cdot C_A$ arises from gluon loops, and $C_A = 3$ is given by the structure of the $SU(3)$ symmetry group. In contrast, the term $2n_f$ arises from fermion loops, and its value depends on the number of color-charged fermions, n_f , which have a mass smaller than Q^2 (the number of so-called *active flavours*). In the Standard Model $n_f \leq 6$, giving

$$\beta_0^{\text{QCD}} \geq \frac{7}{4\pi} > 0, \quad (2.41)$$

which, inserted into Eq. (2.38), implies that the coupling constant of QCD, α_s , decreases asymptotically towards zero for $Q^2 \rightarrow \infty$. In other words, quarks behave as free particles at high energies, a property which is known as *asymptotic freedom*. The coupling constant of QCD becomes large for small momentum transfers and diverges at a scale of approximately 200 MeV. Phenomena mediated by QCD at momentum transfer scales below about 1 GeV, like the structure of hadrons, can only be described using either phenomenological models (see Chapter 3) or using so-called *lattice* QCD calculations in which divergences are avoided through the quantization of space-time [3]. The non-observation of low-energetic, free, color-charged particles in experiments leads to the assumption that such particles cannot be isolated

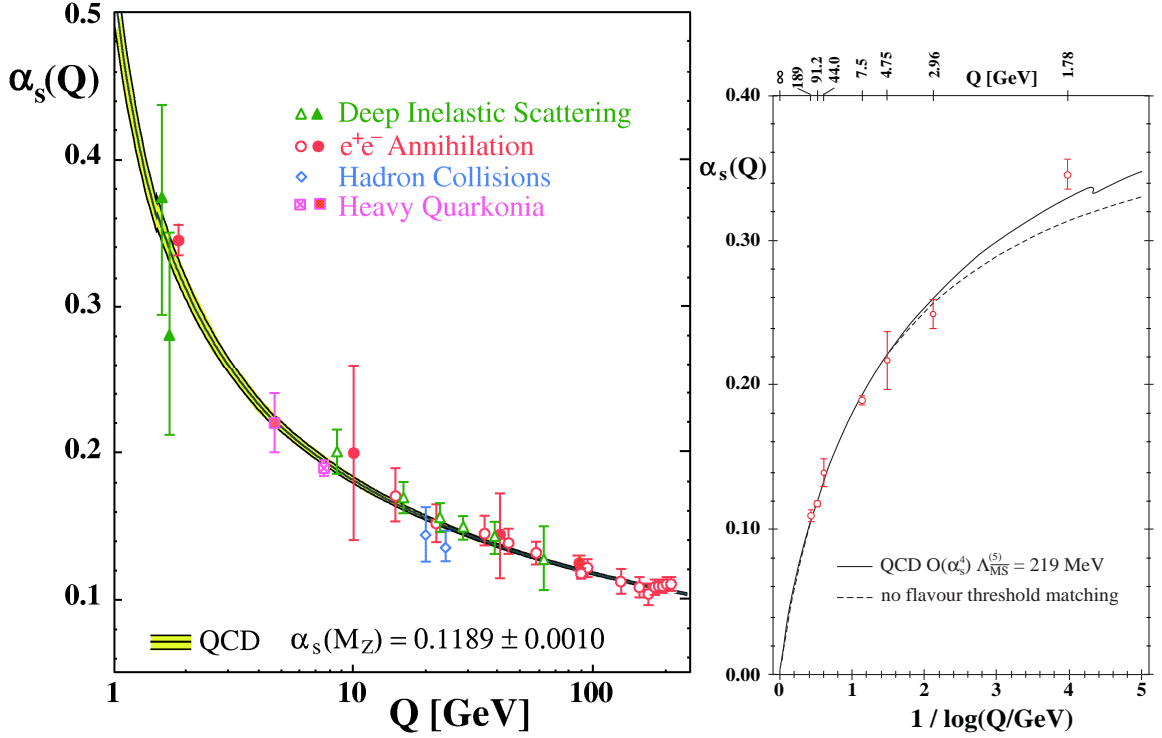


Figure 2.4: The running of α_s as predicted by perturbative QCD calculations is in good agreement with experimental measurements covering a range of momentum transfer scale values of about two orders of magnitude (from Ref. [4]).

singularly, a principle referred to as *confinement*. Comparisons between the running of α_s and experimental measurements at many different energy scales have shown excellent agreement, see Fig. 2.4.

Ten different measurements which are experimentally and theoretically well understood have been evolved from their respective scales to one common scale, taken to be M_Z , and combined to give the current world average of

$$\alpha_s(M_Z) = 0.1189 \pm 0.0010, \quad (2.42)$$

with a χ^2 per degree of freedom of 9.9/9 [4].

When computing a cross section the scale at which the running coupling constant is to be evaluated, μ_R , is not uniquely defined. In principle any positive value can be chosen. In general, a good choice of μ_R is one that minimises the size of higher-order corrections, and frequently a scale characteristic of the hard scattering is used. The uncertainty of a computation due to the non-unique choice of μ_R is called the *renormalization scale uncertainty*, and by convention it is evaluated as the effect of scaling μ_R up and down by a factor of two. The renormalization scale uncertainty is not a Gaussian uncertainty, and it can be interpreted as a measure of the *relative* predictive power of different theory predictions. In general, the more orders in the perturbation series which are included in a calculation, the lower is the dependency of the calculation on the choice of renormalization scale. For the exact, all-order solution the dependency vanishes.

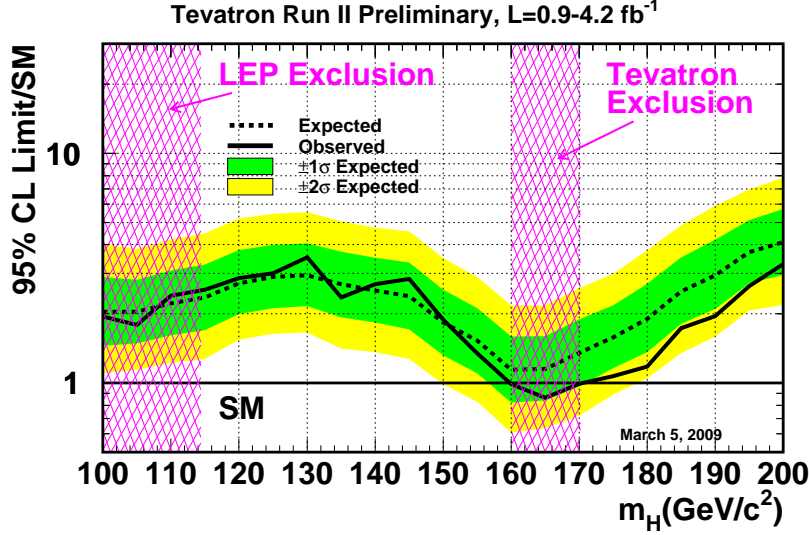


Figure 2.5: Tevatron limits on the Standard Model Higgs production cross section scaled to the Standard Model prediction [6].

The large value of α_s compared with α_{QED} ($\approx 1/137$) means that perturbation series for the matrix elements for QCD processes in general converge more slowly than those for QED processes, making the inclusion of higher-order terms important for producing accurate theoretical predictions. An additional complication when computing QCD cross sections arises from the property of confinement: the observable particles resulting from a QCD interaction are not identical with the particles appearing in the Lagrangian. Phenomenologically motivated models whose parameters are tuned to a large range of experimental measurements are needed to connect incoming and outgoing QCD partons with observable hadrons. How these complications are treated in the calculation of QCD cross sections is described in Chapter 3.

2.5 Detecting the Higgs boson

The Higgs boson is the only Standard Model model particle whose existence has not been confirmed experimentally. Detecting the Higgs boson is one of the main goals of the particle physics community. The mass of the Higgs boson (m_H) is not predicted by the Standard Model, but several experimental and theoretical constraints exist. Unitarity arguments from WW scattering allow an upper limit of $\sim 1 \text{ TeV}$ to be deduced. A lower limit of 114.4 GeV at a confidence level (C.L.) of 95% arise from direct Higgs boson searches at the Large Electron-Positron Collider (LEP) [5]. As shown in Fig. 2.5, searches at the Fermilab Tevatron Collider have excluded a Standard Model Higgs boson with m_H in the range 160–170 GeV at 95% C.L. Fits of Standard Model parameters to electroweak precision measurements [7] are sensitive to the value m_H via virtual corrections, allowing for the extraction of indirect

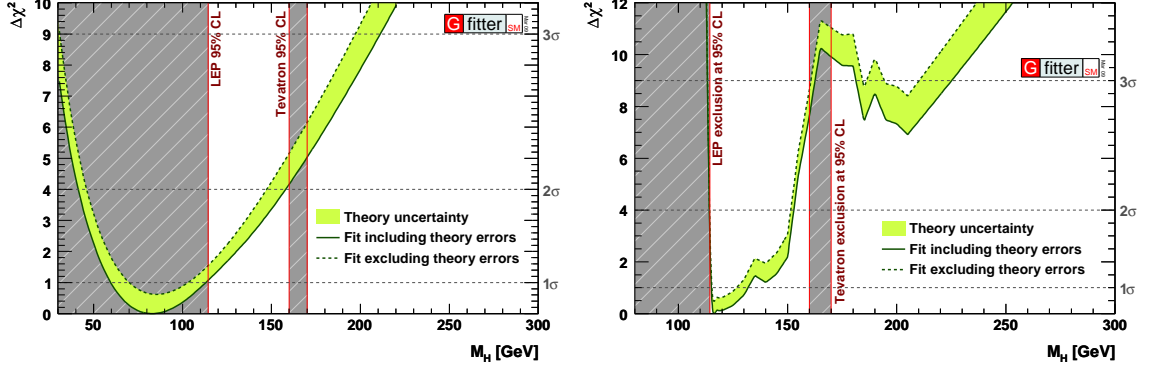


Figure 2.6: $\Delta\chi^2$ versus m_H arising from a fit to electroweak precision measurements (left), and from a fit to electroweak precision measurements combined with direct searches (right) [7].

information on the Higgs boson mass. Current fits to electroweak precision measurements yield $m_H = 83^{+30}_{-23}$ GeV, as shown in Fig. 2.6 (left). Including direct searches from LEP and Tevatron give $m_H = 116.4^{+18.3}_{-1.3}$ GeV, see Fig. 2.6 (right). The 2σ and 3σ intervals for m_H are $[114, 150]$ GeV and $[[113, 163] \cup [182, 226]]$ GeV, respectively (including theory errors). The Standard Model Higgs boson couples to fermions proportionally to $m_f/2m_W$ and to gauge bosons proportionally to the gauge boson mass. The largest contribution to the Higgs cross section comes from gluon fusion via a top-quark loop, followed by the vector-boson fusion channel. The Standard Model Higgs boson preferentially decays into WW or ZZ pairs for the m_H range where this is kinematically allowed, see Fig. 2.7. Below ~ 150 GeV decays into $b\bar{b}$ dominate, with small but important contributions coming from $\tau\tau$ and $\gamma\gamma$. Studies of the Standard Model Higgs boson discovery potential for the LHC experiments have identified $H \rightarrow \gamma\gamma$ via gluon fusion and $Hq\bar{q} \rightarrow \tau\tau q\bar{q}$ via vector-boson fusion as the most promising discovery channels for the low Higgs mass region favored by electroweak precision measurements [8]. Recent studies [9] introduce the associated production of W or Z with a high- p_T $H(\rightarrow b\bar{b})$ as a potentially important discovery channel in case of a Higgs mass below ~ 130 GeV. For a Higgs mass above ~ 140 GeV, $H \rightarrow WW$ and $H \rightarrow ZZ$ show the largest discovery potentials.

2.6 Problems of the Standard Model

The Standard Model has enjoyed an unprecedented level of success since it was introduced during the 1970s. It is in agreement with all existing particle-physics measurements⁴, and it has allowed for several new predictions which have been confirmed by experiments. The only Standard Model particle not detected experimentally so far is the Higgs boson.

In spite of its successes, there are many conceptual problems to which the Standard Model offers no answer. Some of these problems are summarized in the following list:

- Why are particles separated into fermions and bosons?

⁴The only exception being neutrino-oscillation measurements which imply non-zero neutrino masses and therefore the existence of right-handed neutrinos.

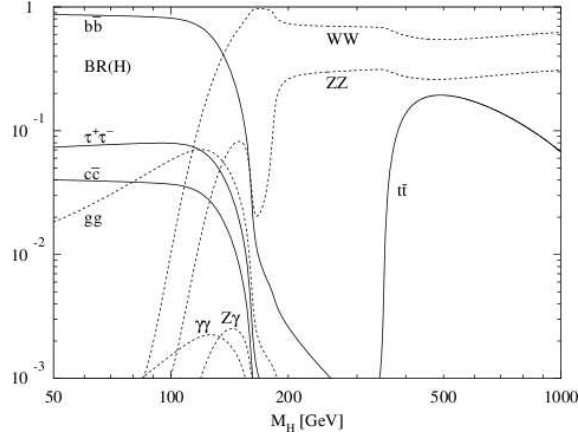


Figure 2.7: The branching ratios of the Standard Model Higgs boson [10].

- Why are fermions separated into leptons and quarks?
- Why are there exactly three different fermion generations?
- Why do the observed fermion masses span over more than ten orders of magnitude, from neutrinos with masses below ~ 3 eV to the top quark with a mass of around 170 GeV?
- The Standard Model allows for a term in \mathcal{L}_{QCD} which is CP violating and proportional to the so-called Θ parameter. Measurements of the neutron electric dipole moment leads to the limit $\Theta < 10^{-9}$ [11]. The Standard Model offers no explanation for why Θ is so close to zero.
- Why is gravity so much weaker than the Standard Model forces?
- All attempts to formulate a quantized description of lead to non-renormalizable theories. The Standard Model ignores all contributions from gravitation. This is acceptable at low energies due to the relative weakness of gravity, but not at the Planck scale ($M_P = 1.2 \times 10^{19}$ GeV) where gravitation is expected to become strong.
- The running coupling constants of the Standard Model lead to the coupling constants of the $U(1)_Y$, $SU(2)_L$ and $SU(3)$ gauge interactions to become *almost* equal at around 10^{13} GeV (see Fig. 2.8, left). The Standard Model fails to offer an explanation for this “almost-unification” of the three forces at high scales.
- In analogy with the running of the coupling constants (see Sect. 2.4), an observable particle mass is a function of a bare mass and higher-order loop contributions. Assuming the Standard Model to be valid up to the Planck scale the quadratically divergent loop contributions to m_H are of the order of M_P^2 . For the observable Higgs mass to be of $O(100 \text{ GeV})$, as indicated by electroweak precision measurements, the bare quadratic Higgs mass must be fine-tuned to an accuracy of 10^{-34} . The large size of the Planck scale compared to the electroweak scale of ~ 100 GeV is referred to as the *hierarchy problem* of the Standard Model.

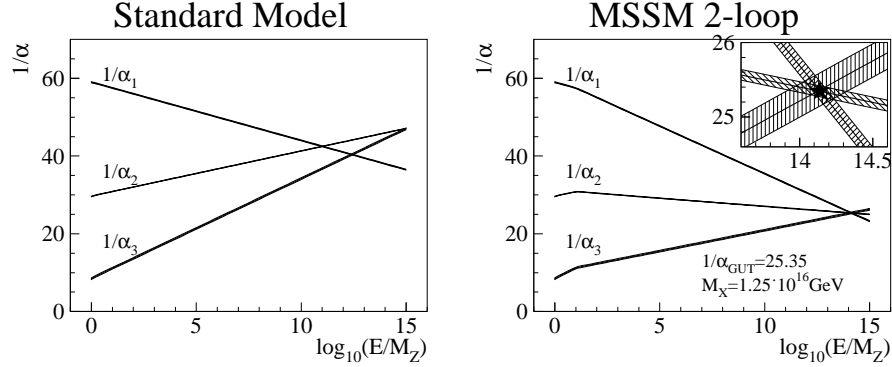


Figure 2.8: Running coupling constants in the SM (left) and in a MSSM scenario (right).

- Cosmic microwave background measurements and other astroparticle-physics measurements have shown that Standard Model particles make up only about 5% of the universe [12]. Dark matter, not part of the Standard Model, make up about 23% of the universe. The remaining 72% is assigned to dark energy which is also not described by the Standard Model.

2.7 Extensions to the Standard Model

Many different extensions to the Standard Model have been suggested to overcome one or more of the problems listed above. Common to these *beyond-the-Standard-Model* (BSM) models is the predicted existence of new particles. Apart from completing the experimental verification of the Standard Model by detecting the Higgs boson, the main goal of present and future particle-physics experiments is the discovery of such BSM particles.

The most promising extensions to the SM are probably *supersymmetric* (SUSY) models which expand the gauge symmetries of the Standard Model with a symmetry between fermions and bosons, leading to each SM fermion (boson) having a boson (fermion) SUSY partner. The non-observation of SUSY partners with masses equal to the Standard Model particles implies that supersymmetry, if realized in nature, has to be broken. Supersymmetry provides solutions to several of the problems of the Standard Model. The introduction of supersymmetry removes the fundamental distinction between fermions and bosons. In case the so-called *R-parity* quantum number is conserved, SUSY also predicts the existence of a stable, lightest SUSY particle which is a candidate for dark matter (see for instance Ref. [13]). The appearance of SUSY particles in loop corrections to Standard Model quantities removes the quadratic divergence of the Standard Model Higgs mass, and also replaces the almost-unification of the Standard Model coupling constants at high energies with a proper unification. No SUSY particle has been detected experimentally so far in spite of many searches being performed in a long list of experiments. Searches for SUSY particles has a high priority at the upcoming LHC experiments.

Motivated by the non-observation of SUSY particles, alternative explanations for the cancellation of the quadratically divergent corrections to the Higgs mass have been proposed. So-called *Little Higgs* models [14] cancel divergent contributions from Standard Model fermions

(bosons) using new fermions (bosons). Whereas SUSY solves the divergence problem through exact cancellations, little-Higgs models only postpones the problem, requiring additional BSM physics, e.g. SUSY, to appear at a higher energy scale. Another strategy for solving the hierarchy is the introduction of large extra dimensions through for example the ADD model [15]. Only gravity is allowed to propagate into the extra dimensions, resulting in a dilution of the strength of gravity compared with the Standard Model forces. This corresponds to a small value of the Planck mass (~ 1 TeV is possible), thereby controlling the quadratic contributions of the Higgs mass.

The Standard Model model can also be extended by unifying the $SU(3)$ and $SU(2) \times U(1)$ symmetries into a larger gauge group. Such *grand-unified theories* (GUT) predict the existence of particles which carry both non-zero color and a non-zero lepton number, referred to as leptoquarks. Another prediction of GUTs is the existence of gauge bosons similar to the Standard Model ones, but with a higher mass. A generic high-mass version of the Z boson is referred to as Z' .

The non-observation of CP violation in the QCD sector has lead to the proposal of a new symmetry which forbids a CP violating term from appearing in \mathcal{L}_{QCD} , leading to a new particle, the axion [16]. Axions are searched for in experiments like CAST [17], exploiting the fact that axions can be converted into photons in the presence of a strong magnetic field.

Chapter 3

Phenomenology of $p\bar{p}$ Collisions

This Chapter gives an overview of various techniques and models commonly applied for deriving observable predictions from the Lagrangian of the Standard Model. The production of Z/γ^* in $p\bar{p}$ collisions is used as an example, but most statements are also valid in the context of any Standard Model process.

Using perturbation theory, transition probabilities can be expressed in terms of a power-series expansion in the coupling constants of the forces involved in the interaction. The initial and final-state particles are quanta of the fields appearing in the Lagrangian. For example, the cross section for $q\bar{q} \rightarrow e^+e^-$, including higher-order terms only¹ in α_s , can be written as

$$\sigma = \alpha_{\text{EW}}^2 \sum_{i=0}^{\infty} \alpha_s^i A_i, \quad (3.1)$$

where the coefficients A_i are functions of the four-momentum vectors of the initial and final-state particles. The term corresponding to $i = 0$ is referred to as the *leading-order* (LO) term, the $i = 1$ term as the *next-to-leading-order* (NLO) term, or in general, the $i = m$ term is the $N^m\text{LO}$ term. Higher-order terms can be grouped into two categories: real-emission terms, corresponding to Feynman diagrams with an additional final-state particle compared with the LO diagrams, and virtual-correction terms, arising from the interference between diagrams containing an internal loop and LO diagrams (see Fig. 3.1). With current techniques it is not possible to take all terms in Eq. (3.1) into account. Instead, one has to identify and take into account the most important terms, ignoring all other terms. As illustrated below, the relative importance of terms depends on the observable. A simplified overview which terms are taken into account in the various models described in this Chapter is given in Tab. 3.1 at the end of the Chapter.

¹Motivated by higher-order terms in α_s being more important than those in α_{EW} for the study presented in this thesis.

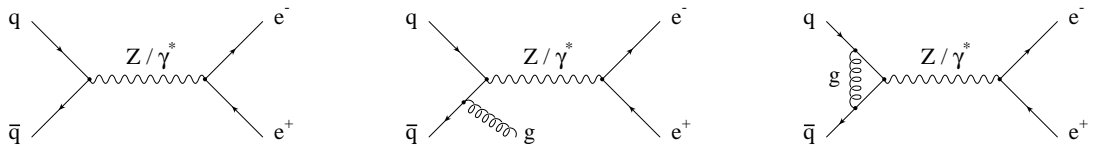


Figure 3.1: Feynman diagrams for $q\bar{q} \rightarrow e^+e^-$ production: LO (left), real-emission NLO (center) and virtual-correction NLO (right).

3.1 Fixed-Order Calculations

Assuming that $\alpha_s \ll 1$, and that the coefficients A_i do not play a deciding role in the ordering of the terms in Eq. (3.1) – referred to as the *fixed-order assumption* below – the term $\alpha_s^i A_i$ is significantly smaller than $\alpha_s^{i-1} A_{i-1}$. In this case a good approximation to the cross section is found by including only the n first terms, where n is as large as technically possible. The resulting calculation is referred to as a *fixed-order calculation*. Real-emission terms are divergent in the limit where a parton is soft or collinear with respect to another parton. These divergences cancel against opposite-sign divergences in virtual-correction terms, yielding a finite, physical result if all terms to a fixed order are taken into account [18].

Current state-of-the-art fixed-order calculations for 2 partons $\rightarrow Z$ include NNLO order QCD terms [19]. For $q\bar{q} \rightarrow Z + N$ partons, calculations with NLO QCD terms exist for $N = 1, 2$ [20]. Calculations for $W + 3$ partons production which includes the most important terms at NLO (*leading-color approximation*) have recently been presented [21, 22, 23], and the corresponding calculation for $Z + 3$ partons is expected to follow soon. If one ignores virtual corrections, the calculation for $q\bar{q} \rightarrow Z/\gamma^* \rightarrow e^+e^- + N$ partons is automatised for all N [24, 25], but computer power constraints currently limit N to ~ 6 [25].

3.2 Including Terms to all Orders

If a final-state parton is soft or collinear with respect to another parton, the so-called *soft/collinear limit*, a fixed-order calculation does not include all large terms. In general, the coefficients A_i can be written as

$$A_i = \sum_{k=0}^{2i} B_{ij} L^j. \quad (3.2)$$

The function L becomes large in the soft/collinear limit due to logarithmic factors like

$$\log \frac{Q^2}{q^2}, \quad (3.3)$$

with Q^2 being a high momentum scale characteristic for the interaction, e.g. M_Z for Z -production, and q^2 being the transverse momentum of one parton with respect to another. For $q^2 \rightarrow 0$, L becomes large and A_i is said to be *logarithmically enhanced*. For such soft and/or collinear final-state constellations, the suppression of higher order terms due to the coupling constants is counter-acted by large logarithmic factors. As a result, higher-order terms might be larger than lower-order terms, and the fixed-order approach breaks down. An example where logarithmically enhanced terms are important is the low- p_T region of the p_T^{Z/γ^*} distribution in $p\bar{p} \rightarrow Z/\gamma^* \rightarrow e^+e^-$ production. As seen in Fig. 3.2, fixed-order calculations at NLO and NNLO are unable to describe data below ~ 5 GeV.

3.2.1 Analytical Resummation

The break-down of the fixed-order approximation in the soft/collinear limit does not reflect a fundamental problem with the perturbative approach, only that one has failed to account for the most important terms in the perturbative expansion. The solution is to re-order the

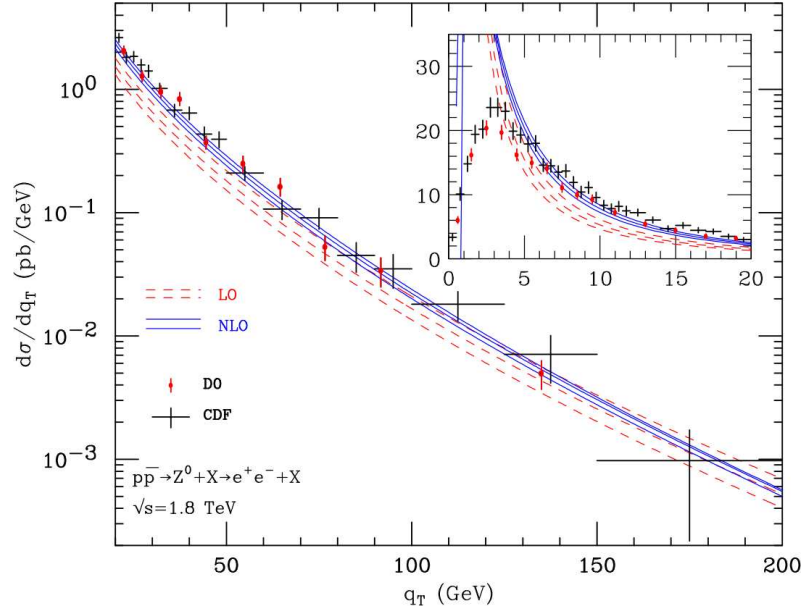


Figure 3.2: Comparisons between measurements of p_T^Z in $p\bar{p} \rightarrow Z/\gamma^* \rightarrow e^+e^-$ from the CDF and DØ experiments and fixed-order calculations including terms of order up to α_s^1 , labelled LO, and up to α_s^2 , labelled NLO [26]. Each prediction is shown for a scale choice of M_Z^2 , as well as for $0.5 \cdot M_Z^2$ and $2 \cdot M_Z^2$.

terms in Eq. (3.1) from being a power-series in the coupling constants, to being a power-series in the coupling constants times a logarithm, L . Terms proportional to $\alpha_s^i L^{2i}$ are said to be the *leading log* (LL) terms, and $\alpha_s^i L^{2i-1}$ the *next-to-leading log* (NLL) terms and so on. A calculation which takes all LL terms into account is referred to as a LL *resummation* calculation. Analytical resummation is based on identifying patterns in the coefficients B_{ij} in Eq. (3.2) which allows the use of relations like $\sum_{i=0}^{\infty} (x^i/n!) = e^x$ to account for the most important logarithmically enhanced terms to all orders in α_s . Analytical resummation of p_T^{Z/γ^*} in $p\bar{p} \rightarrow Z/\gamma^* \rightarrow e^+e^-$ production are able to describe the low- p_T region where the fixed-order calculation breaks down. The assumption of a resummation calculation, that L is large, is not true for final-states with hard, well-separated partons, corresponding to large values of p_T^Z . A disadvantage of analytical resummation is that the technical details of the calculation depend on the observable, not just on the process.

3.2.2 Parton Shower

Parton-shower algorithms represent an alternative approach to analytical resummation for including logarithmic terms to all orders in the coupling constants. The starting point of a parton-shower algorithm is a set of n final-state partons (*matrix-element partons*) generated according to a $2 \rightarrow n$ matrix-element calculation. As for resummation, the leading logarithmic terms correspond to diagrams which contain additional partons which are soft and/or collinear to one of the matrix-element partons. In a full matrix-element calculation the evaluation of $|\mathcal{M}|^2 = |\mathcal{M}_1 + \mathcal{M}_2 + \dots|^2$ includes interference terms, e.g. $\mathcal{M}_1 \cdot \bar{\mathcal{M}}_2$, corresponding to the interference between an additional parton being emitted off two different matrix-element

partons, each possibility being described by a different matrix-element term, \mathcal{M}_1 and \mathcal{M}_2 . Parton-shower algorithms are based on the observation that the contributions of interference terms vanish in the soft/collinear limit, meaning that $|\mathcal{M}|^2 \approx |\mathcal{M}_1|^2 + |\mathcal{M}_2|^2 + \dots$. In this limit one can adopt the picture where a parton is emitted from another parton in a $1 \rightarrow 2$ *splitting*. The functions which describe the probabilities and kinematics of such $1 \rightarrow 2$ splittings are referred to as *Altarelli-Parisi splitting functions* [27]. A parton-shower algorithm generates emissions which are ordered from large to small values of an *evolution variable* which can for instance be the virtuality of the incoming parton (Q^2 *ordering*), the relative transverse momentum of the two partons (p_T *ordering*), or the angle between the two partons (*angular ordering*). The parton-shower algorithm starts from a maximal value of the evolution variable, the *starting scale*, and terminates at a minimal value, the *cut-off scale*. The interval between the starting and cut-off scales is divided into many small intervals, and the probability that no emission takes place in each interval is given by so-called *Sudakov factors* [28]. Based on a random-number generator, decisions are taken on whether or not an emission takes place in each interval, and, if it does, on the resulting kinematics of the $1 \rightarrow 2$ splitting. The value of the starting scale is not uniquely defined, and which value to use is normally guided by assuming that it should be a scale which is characteristic for the process. As an example, for Q^2 -ordered showers the scale of the hard scattering is typically used, meaning M_Z^2 for $q\bar{q} \rightarrow Z/\gamma^* \rightarrow e^+e^-$ production. The lower cut-off scale must be set to a value large enough that QCD is still perturbative, and it is typically defined by the scale at which the model which translates partons into hadrons (*hadronization model*, see below) was tuned, typically 1 – 2 GeV. The splitting probability depends on α_s and the renormalization scale at which its numerical value is evaluated. Theoretical arguments favour a scale proportional to the relative p_T of the splitting [29]. Starting from a core $2 \rightarrow n$ process described by a matrix-element calculation, one distinguishes between an *initial-state* parton-shower algorithm which is applied to the two incoming partons, and a *final-state* algorithm which is applied to outgoing partons. Whereas final-state algorithms trace partons forwards in time, initial-state algorithms go backwards in time from the hard scattering, taking into account additional constraints arising from the fact that the two initial partons are parts of protons/antiprotons [30].

Parton-shower algorithms are closely related to analytical resummation in the sense that both aim to include the most important logarithmic terms to all orders in α_s . Whereas resummation gives a transparent mathematical calculation in the sense that it is clear which terms in the perturbative power-series expansion are accounted for, it is less obvious which terms are actually included by a parton-shower algorithm. Currently available algorithms include all leading-log terms. Four-momentum conservation, which is violated in a LL calculation, is restored through a separate step after the parton shower has terminated by modifying the kinematics of the final-state partons. One advantage of parton-shower algorithms is that they produce fully exclusive and differential predictions². In addition, using a parton-shower algorithm allows for predictions about stable, observable particles instead of about partons through the use of a hadronization model. Particle-level predictions are needed to perform a simulation of the performance of the experimental detector setup. Such detector simulations are essential to achieve a correct interpretation of experimental data, see Chapt. 4.

While the parton-shower approach is only theoretically justified in the soft/collinear limit, it

²In the sense that, whereas a resummed calculation will typically predict one single distribution, like p_T^Z in $q\bar{q} \rightarrow Z/\gamma^* \rightarrow e^+e^-$, a parton-shower approach will allow for the generation of the full structure of single events, allowing for the study of all kinematic properties of the event, like p_T^Z , p_T^{jet} , $\Delta\phi(Z, \text{jet})$ and so on.

is frequently used to also generate hard, well-separated partons, motivated by the assumption that it is better to have some emissions in this region than none at all: if one has some emissions one can reweight the events to correct the theoretically spurious parton-shower predictions to reproduce either more well-founded theory predictions, or experimental data.

3.3 Combining Fixed-Order and All-Order Calculations

So far we have considered calculations which include either all terms to a certain order in the coupling constant or calculations which include the leading logarithmic terms for each order in the coupling constant. Since fixed-order and all-order methods are complementary by construction, an improved prediction can be achieved by combining the two.

3.3.1 Combining Fixed-Order and Analytical Resummation Calculations

As seen in Fig. 3.2 for p_T^Z in $q\bar{q} \rightarrow Z/\gamma^* \rightarrow e^+e^-$ events, a fixed-order calculation includes the most relevant terms at high p_T^Z , but diverges at low p_T^Z . In contrast, a resummed calculation describes data at low p_T^Z , but not at high p_T^Z . There is an overlap between the terms accounted for by the two methods, e.g. both NLO and NLL include the same logarithmic order terms for order α_s . When constructing a combined prediction, all terms taken into account by both methods must be explicitly subtracted once to avoid the problem referred to as *double counting*. To identify such subtraction terms, the resummed calculation must be expanded in terms of α_s and compared with the fixed-order calculation. The combined calculation includes all terms to a certain order in the coupling constant, as well as logarithmic terms to all orders in α_s . Figure 3.3 shows a comparison between experimental data and a calculation combining NLO³ and NLL terms for p_T^Z in $p\bar{p} \rightarrow Z/\gamma^* \rightarrow e^+e^-$ events.

3.3.2 Combining Fixed-Order Calculations and Parton-Shower Algorithms

As mentioned in Sect. 3.2.2, it is less clear which terms are actually taken into account by a parton-shower algorithm than in a resummed calculation. Consequently, the identification of subtraction terms to avoid including a term twice (*double counting*) is more complicated when combining a fixed-order calculation with a parton-shower algorithm than with an analytical resummation calculation. In the case of an electroweak core process like $q\bar{q} \rightarrow Z$ being combined with a QCD parton-shower algorithm, double-counting is avoided trivially – it only becomes an issue if one or more QCD vertices are present in the Feynman diagrams included in the matrix-element calculation, e.g. for $q\bar{q} \rightarrow Zg$.

The first step towards combining parton-shower algorithms with matrix-element calculations for the associated production of a gauge boson with one or more jets was algorithms which reweight the leading initial-state parton-shower emission to reproduce the p_T spectrum predicted by a LO Z +parton fixed-order calculation [31, 32, 33]. This can be achieved by increasing the starting-scale of the parton-shower algorithm until it generates more well-separated partons with large values of p_T than predicted by the fixed-order calculation. Then, each leading parton-shower emission is kept or rejected based on a weight defined by the ratio of the p_T spectrum of the leading parton predicted by the fixed-order calculation to that

³I.e. including terms of order up to α_s^1 . In the present text this is referred to as NLO, whereas in Ref. [26] the term LO is used since this is the first order in perturbation theory where the Z boson has a non-zero p_T .

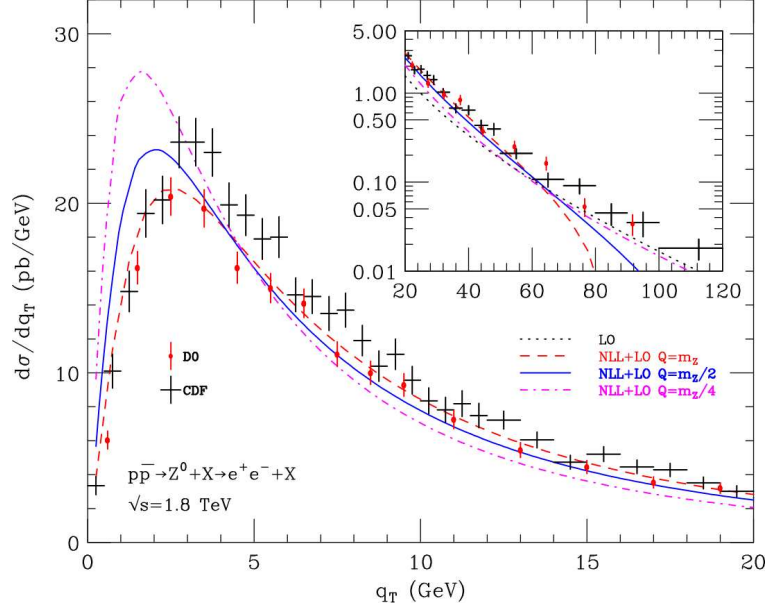


Figure 3.3: Same data-points as in Fig. 3.2, compared with a calculation combining NLL resummation with all terms to order up α_s^1 (labeled NLL+LO), for various choices of the resummation scale, Q , which is a parameter of the resummed calculation, in analogy with the factorization scale [26].

predicted by the artificially enhanced parton-shower algorithm. This technique is referred to as *matching* a parton-shower algorithm to a matrix-element calculation.

Since parton-shower algorithms and matrix-element calculations are formulated in rather different theoretical frameworks, a more formal combination of the two methods than a reweighting of the leading emission requires either that the parton-shower algorithm is formulated in “matrix-element language”, or that the final-state of the matrix-element calculation is translated into “parton-shower language”. An example of the former strategy are algorithms combining matrix-element calculations with a parton shower through a subtraction term technique similar to those used for combining fixed-order and analytical resummation calculations in the MC@NLO [34] and POWHEG [35] event generators. The NLO terms which are taken into account also by the parton shower must be identified and subtracted to avoid double counting. These subtraction terms depend both on the details of the parton-shower algorithm and on the process. In the case of inclusive $Z/\gamma^* \rightarrow e^+e^-$ production the resulting calculation includes all LO and NLO terms in α_s , as well as all LL and some NLL terms to all orders in α_s . This means that the properties of the Z/γ^* system (leading jet) are predicted to next-to-leading (leading) order in α_s , whereas all subleading jets are generated according to the soft/collinear approximation of the parton-shower algorithm. Achieving an improved description for sub-leading partons within this model would require extending the method to merge a NNLO calculation and a parton shower, and this has so far not been done.

The opposite strategy, translating the matrix-element prediction into an input state accepted by a parton-shower algorithm allows for the combination of tree-level LO matrix-elements with parton showers in a process independent way. Such techniques take real-emission terms up to order α_s^N into account, with typically $N \leq 6$, as well as all LL and some

NLL terms to all orders in α_s . Virtual corrections (loop diagrams) are not included, and consequently the matrix-element calculation diverges for soft/collinear partons. These divergences are avoided by requiring all partons to be hard and well-separated. The phase-space region thus ignored is taken into account using a parton-shower algorithm. In the following, the properties of the two main algorithms of this kind are outlined, using $q\bar{q} \rightarrow Z/\gamma^* \rightarrow e^+e^-$ production as an example.

CKKW Merging

The CKKW algorithm [36, 37, 38] starts by generating parton+parton $\rightarrow m+n$ events, where n (m) denote the number of final-state particles carrying (not carrying) a QCD color charge. I.e. for the process $q\bar{q} \rightarrow e^+e^- + 1$ parton, $m = 2$ and $n = 1$. The kinematics of the generated events are given by a tree-level matrix-element calculation. The probability for generating each QCD final-state multiplicity $n = 0, 1, \dots, N$ is given by $\sigma_n / \sum_{i=0}^N \sigma_i$, where the cross sections σ_i are evaluated for a fixed number of m and N is the upper number of QCD partons which are included in the matrix-element calculation. The cross sections are evaluated using a fixed value of μ_F and μ_R which is characteristic for the hard process. Divergences due to the missing virtual corrections are avoided using a phase-space cut requiring each parton to be hard and well-separated with respect to all other partons, thus dividing the final-state phase space into one region described by the matrix-element calculation, and one region explicitly ignored by it. The definition of hard and well-separated is that each pair of partons is more distant from each other in phase space than a value Q_{cut}^2 , with the measure of distance being given as the relative transverse momentum of the two partons, following the k_T jet algorithm [39].

The next step is to reconstruct the internal state of a parton shower which, if it was “paused” after the appropriate number of splittings, would have generated the matrix-element final state. For the example of a parton+parton $\rightarrow e^+e^- + 1$ parton event this amounts to assuming that the final-state parton was emitted from the initial-state parton to which it is closest in phase space, with the parton-shower evolution variable scale at which the emission took place being given by the phase-space distance. A Sudakov weight accounts for the parton shower not having emitted additional partons at a higher evolution-variable scale. For each parton-shower emission, α_s is evaluated at the scale given by the relative p_T of the two partons after the splitting, whereas for the matrix-element calculation α_s is evaluated at the scale of the hard interaction. To compensate for this mismatch, the event is given a weight equal to the ratio of the value of α_s evaluated at the former scale to the value of α_s evaluated at the latter scale. For initial-state emissions, an additional reweighting factor is applied to account for the dependence of the parton shower on the proton structure. This finishes the reconstruction of the history of a parton shower which, being paused in the middle, would have generated the n matrix-element final-state. The parton shower is then “restarted” to account for the soft/collinear part of the phase space which was explicitly left out of the matrix-element calculation. Double counting is avoided by vetoing parton-shower splittings which would generate two partons which have a larger phase-space distance to each other than Q_{cut}^2 , giving a so-called *vetoed* parton shower.

MLM Merging

The MLM algorithm [40, 41] for merging tree-level matrix-element calculations and parton showers is an alternative to the CKKW algorithm. The main difference between the two is how

they prevent double counting. Whereas CKKW vetoes single parton-shower emissions, MLM does not interfere with the parton-shower algorithm directly. Instead, events are rejected if the number of hard, well-separated partons after the parton shower has finished is not identical to the number of partons included in the matrix-element calculation. Whereas CKKW uses the k_T measure of distance to separate the matrix-element and parton-shower parts of phase space, the MLM algorithm uses a cut on the p_T of $\Delta R = \sqrt{\Delta\phi^2 + \Delta y^2}$ cone jets.

3.4 From Partons to Hadrons

So far we have dealt with partonic processes where the initial-state and final-state objects were quanta of the fields appearing in the Standard Model Lagrangian. Due to the confining property of the strong interaction (see Sect. 2.4), quarks and gluons are not observable objects and only offer a high- p_T description of QCD processes. For the example of a $p\bar{p}$ collider, the partonic process $q\bar{q} \rightarrow e^+e^-$ is only a high- p_T model of the physical $p\bar{p} \rightarrow e^+e^-$ process. The observable cross section, σ , can be related to the high- p_T partonic cross section, $\hat{\sigma}$, via a factor \mathcal{N} , into which all non-perturbative, low- p_T effects are absorbed:

$$\sigma = \hat{\sigma} \otimes \mathcal{N}. \quad (3.4)$$

Since \mathcal{N} is not predicted by perturbation theory, this equation is only useful due to a property of QCD referred to as factorization, implying that \mathcal{N} is independent of the details of the partonic process. Consequently, \mathcal{N} can be determined from $\hat{\sigma}_1$ and the measured σ_1 for a set of processes, and then used together with $\hat{\sigma}_2$ for a different process to predict the observable σ_2 . The scale which separates effects being accounted for in $\hat{\sigma}$ and \mathcal{N} , i.e. which defines the boundary between “high- p_T ” and “low- p_T ”, is referred to as the *factorization* scale. The factorization scale is an unphysical artifact of the calculation, and the dependence on this scale would cancel exactly for the exact solution. In general, the more orders in the coupling constant which is included in a calculation, the lower is the factorization scale dependency.

3.4.1 Hadronization

The concept of PDFs is introduced to connect incoming, observable hadrons with the incoming partons of the perturbative calculation. A similar picture is used for connecting the outgoing partons of the calculation with outgoing, observable hadrons. This process is referred to as *hadronization*, and it can be illustrated in the process $e^+e^- \rightarrow \text{hadrons}$, described at the parton level as $e^+e^- \rightarrow q\bar{q}$. The simplest hadronization model assumes that the transition from partons to hadrons can be described for each parton independently (*independent hadronization*). The hadronization function $f_p^h(z, k_T^2)$ for a parton of flavor p specifies the probability for the production of a hadron of type h , carrying a fraction $0 < z < 1$ of the parton energy, with transverse energy k_T relative to the parton. Assuming the validity of a partonic $e^+e^- \rightarrow Z/\gamma^* \rightarrow q\bar{q}$ prediction, the parameters of the hadronization functions can be determined so that this prediction is compatible with experimental measurements of the hadrons in $e^+e^- \rightarrow Z/\gamma^* \rightarrow \text{hadron production}$. Independent hadronization models were not able to describe the topology of $e^+e^- \rightarrow Z/\gamma^* \rightarrow q\bar{q}g$ events seen by the JADE experiment [42]: the hadron-density is enhanced in the region between a quark and a gluon, and such correlations cannot be accounted for by independent hadronization. However, such

correlations were predicted by the *Lund string model* [43] which assumes that the hadronization process is a function of the color field, or string, spanned out between pairs of quarks. Gluons are represented as kinks along the string. Each string is independently decayed into hadrons using hadronization functions. Another model able to describe experimental data, the *cluster hadronization model* [44], is based on the independent hadronization of clusters formed from pairs of neighbouring quarks after all gluons have been forced to split into $q\bar{q}$ pairs.

Every hadronization model depends on the details of the partonic calculations used when extracting the model parameters from experimental data. If $2 \rightarrow 2$ parton-level predictions are used, the effects of higher-order and logarithmically enhanced perturbative terms are absorbed into the hadronization model, leading to the hadronization model becoming dependent on the details of the hard scattering. To assure a process-independent hadronization model, and since perturbative effects can be accounted for more correctly using a parton shower, hadronization models are instead defined using a full matrix-element plus parton-shower prediction for the perturbative part. The fragmentation model will depend on the lower cut-off scale of the parton shower, i.e. on the scale at which the fragmentation model was tuned, typically ~ 1 GeV. This scale is the factorization scale which separates the effects accounted for by the parton shower from those absorbed into the hadronization model. Since hadronization models are only defined for partons at a fixed, low energy scale, they cannot be used to translate the high-energy-scale final-state partons of a fixed-order calculation into hadrons. Doing so would ignore the impact of logarithmically enhanced higher-order terms, invalidating the resulting particle-level prediction.

3.4.2 Parton Distribution Functions

Whereas hadronization models describe the transition from outgoing partons into outgoing hadrons, parton distribution functions (PDFs) describe the relationship between incoming partons and incoming hadrons. The flavors of the incoming hadrons are given by the experimental accelerator setup. A parton distribution function, $f_p^h(x, \mu_F^2)$, describes the probability of finding a parton of flavor p inside a hadron h with the parton carrying a fraction $0 < x < 1$ of the total momentum of the hadron. The PDFs depend on the energy scale at which the hadron is probed, the factorization scale μ_F^2 . This factorization scale separates effects which are accounted for in a matrix-element calculation from those absorbed into the PDFs. The choice of factorization scale is not uniquely defined, giving rise to a systematic uncertainty referred to as the *factorization scale uncertainty*. The observable cross section for $p\bar{p} \rightarrow Z/\gamma^* \rightarrow e^+e^-$ production, σ , can be expressed in terms of the partonic cross section for $q\bar{q} \rightarrow Z/\gamma^* \rightarrow e^+e^-$ production, $\hat{\sigma}$, and the proton/antiproton PDFs, $f_p^h(x, \mu_F^2)$, as

$$\sigma = \sum_{q,\bar{q}} \int dx_1 \int dx_2 \left[f_q^p(x_1, \mu_F^2) f_{\bar{q}}^{\bar{p}}(x_2, \mu_F^2) + f_{\bar{q}}^p(x_1, \mu_F^2) f_q^{\bar{p}}(x_2, \mu_F^2) \right] \hat{\sigma}, \quad (3.5)$$

where the sum runs over all possible initial-state quark flavors. Equation (3.5) can be used either to determine the PDFs if σ and $\hat{\sigma}$ are known, or to predict the observable σ from the perturbative calculation of $\hat{\sigma}$ if the PDFs are known. Due to the property of universality PDFs extracted by fits to experimental measurements for one class of processes can be used to make observable predictions for all other processes. Extracting PDFs from data using Eq. (3.5) introduces a dependence on the perturbative order used for computing $\hat{\sigma}$ and for defining the running of α_s . Several different sets of PDFs are available, for instance those determined

by the CTEQ [45] and MSTW [46] collaborations. As seen above, hadronization model parameters are tuned to data for partons at a fixed energy scale from parton-shower-based parton-level predictions. Consequently, hadronization models cannot be used for fixed-order calculations, and such calculations are only able to make predictions at the parton level. In contrast, the energy-scale dependence of the PDFs can be described perturbatively using *DGLAP* equations⁴ [27, 47, 48]. Consequently, PDFs can be defined independently of parton showers, and they can be used both in analytical calculations and in parton-shower based event generators. The x -dependence of the PDFs is not known theoretically and must be extracted from experimental measurements. The PDFs of the proton/antiproton have been determined over an x range of approximately $[10^{-5}, 1]$ [45] by fitting perturbatively calculated differential cross section predictions to a long range of experimental measurements. The measurements used come from $ep \rightarrow eX$ deep-inelastic scattering experiments from the HERA experiments, QCD multijet measurements from the Tevatron, asymmetry measurements in vector-boson production from the Tevatron, as well as a range of lower-energy fixed-target measurements [49]. Once determined by fitting to data, the PDFs can be evolved in μ_F^2 and used together with perturbative calculations and Eq. (3.5) to predict observable cross sections, both in the context of analytical and parton-shower based calculations.

3.4.3 Intrinsic p_T

It is assumed that partons restricted inside a hadron will have a non-zero, intrinsic transverse momentum due to Heisenberg's uncertainty principle. This is a non-perturbative phenomenon that is not accounted for in PDFs, and is normally modelled by assigning a non-zero p_T to the two initial partons according to a Gaussian function whose width is extracted from measurements of the low- p_T^Z region for $p\bar{p} \rightarrow Z/\gamma^*$ production. While theory suggests an intrinsic p_T of a few hundred MeV, parton-shower based calculations have been found to favour values as high as 2 GeV [50, 51]. This discrepancy might indicate that the impact of perturbative, sub-leading logarithmic terms not accounted for in the LL parton shower are absorbed into the intrinsic p_T model [26].

3.5 Underlying Event

The term *underlying event* is not well-defined in the literature. Here, we use it to refer to any interaction occurring during a single bunch-crossing (see Chapter 4) which is not taken into account in the description of the hard $2 \rightarrow n$ scattering. This definition includes secondary interactions between the $p\bar{p}$ pair of the hard scattering, the description of the remnants of the $p\bar{p}$ pair, and of interactions between other $p\bar{p}$ pairs in the same bunch crossing.

3.5.1 Multiple Parton Interactions

Multiple parton interaction (MPI) models aim at accounting for the possibility that more than one parton-parton interaction occurs for the $p\bar{p}$ pair of the hard scattering. The existence of such secondary interactions was established in the $\gamma + 3$ jets channel by selecting events where two and two objects (i.e. γ +jet and jet+jet) have similar values of p_T and are back-to-back in the transverse plane [52]. The amount of such event exceeded what could be

⁴Based on partonic $1 \rightarrow 2$ splitting processes described by the same Altarelli-Parisi splitting functions [27] which are used in parton-shower algorithms.

explained by perturbative calculations of parton+parton $\rightarrow \gamma + 3$ jets. A widely used MPI model is one [53] based on a series of kinematically independent, perturbative parton-parton scatterings arising from a single $p\bar{p}$ pair. Having generated the hard scattering the PDFs are modified to reflect the momentum and color carried by the two hard scattering initial-state partons. Secondary interactions, which are assumed to be pure QCD processes⁵, are generated according to perturbative 2 \rightarrow 2 calculations. A lower p_T cut-off for the secondary interactions is used to avoid the low- p_T divergences of the 2 \rightarrow 2 QCD cross section and to tune the predictions to data [54]. Assumptions are made on the radial distribution of partons inside the protons, and each collision is assigned an impact parameter defining the “overlap” between the $p\bar{p}$ pair. The probability of multiple interactions increases with increasing overlap.

3.5.2 Beam Remnants

Having extracted one or more initial-state parton to account for the hard scattering as well as any multiple parton interaction scatterings (referred to as the *drawn partons* below), the color-charged remnant of the incoming (anti)proton needs to be modelled. Depending on the flavours of the extracted partons, the remnant is described by a di-quark in case the initial-state particle of the perturbative calculations is a quark, or a quark + di-quark constellation in case of a gluon. The color charge of the remnant is given by the color of the drawn partons, and by the fact that the incoming (anti)proton was color neutral. The longitudinal momentum component of the remnant is defined as the fraction of the original (anti)proton momentum not carried away by the drawn partons, whereas the transverse component is given as the opposite of the intrinsic p_T assigned to the drawn partons. In general, $p_T^{\text{remnant}} \ll p_z^{\text{remnant}}$, and the beam remnants do not give an important contribution to high- p_T observables.

3.5.3 Multiple $p\bar{p}$ Interactions

Due to the experimental setup (see Chapt. 4), there is a possibility that more than one $p\bar{p}$ interaction takes place during each bunch crossing, and such additional interactions are referred to as *pile-up events*. Secondary $p\bar{p}$ collisions can either be accounted for using model predictions [50], or by overlying the simulated hard scattering process with detector signals registered during an arbitrary bunch crossing. The latter method was used for this study, as is described in more detail in Sect. 4.3.

3.6 Event Generators

The term *event generator* is used for a computer code which uses a Monte-Carlo technique to generate events consisting of a list of initial-state and final-state objects with specified four-momentum vectors. The events are generated such that if a kinematic distribution is plotted for a large sample of events, it will reproduce the distribution predicted by the calculation on which the event generator is based. The events can either be *weighted*, i.e. each event has a weight which must be used when filling the histogram of an observable, or *unweighted*, corresponding to each event having a weight of 1. An event generator is said to be a *parton-level* (*particle-level*) event generator if the list of initial-state and final-state objects consist of partons (observable particles).

⁵Justified by the large cross section of pure QCD processes compared to EW processes.

3.6.1 MCFM

MCFM [55] is a FORTRAN-based parton-level event generator based on fixed-order calculations at LO and NLO for a long list of processes. For $Z/\gamma^* + N$ partons it delivers predictions at NLO for $N \leq 2$ and LO for $N \leq 3$.

3.6.2 PYTHIA

PYTHIA [50] is a FORTRAN-based particle-level event generator based on LO matrix-element calculations and, depending on version and configuration, either a Q^2 -ordered or a p_T -ordered parton-shower algorithm. For Z/γ^* production the first parton-shower emission is reweighted to a LO $Z/\gamma^* + 1$ parton matrix-element calculation. Hadronization is accounted for using the Lund string model. Depending on the choice of parton-shower algorithm there are two different multiple parton interaction models. In the model used together with the Q^2 parton shower, described in Sect. 3.5.1, the MPI partons are not passed on to a parton shower, but this has been implemented in the newer model used together with the p_T -ordered shower. Intrinsic p_T is described by a Gaussian model.

3.6.3 HERWIG and JIMMY

HERWIG [56] is a FORTRAN-based particle-level event generator which is similar to PYTHIA. It uses an angular-ordered parton shower and a cluster-hadronization model. For Z/γ^* production the first parton-shower emission is reweighted to LO $Z/\gamma^* + 1$ parton matrix-element calculations, using two different reweighting schemes for the low and high p_T regions. HERWIG has its own underlying event model, but a more detailed description of multiple parton interaction is offered by connecting HERWIG to the JIMMY [57] MPI generator.

3.6.4 SHERPA

SHERPA is a C++-based event generator which uses the CKKW algorithm (see Sect. 3.3.2) to merge tree-level matrix-element predictions with a virtuality-ordered parton-shower algorithm. Partons are translated into hadrons using a cluster-hadronization model. The multiple parton interaction model is similar to the one described in Sect. 3.5.1, but extended to apply parton showering to the $2 \rightarrow 2$ MPI processes.

3.6.5 ALPGEN+PYTHIA and ALPGEN+HERWIG+JIMMY

ALPGEN+X is a FORTRAN-based code which combines tree-level matrix-element predictions from ALPGEN with the PYTHIA or HERWIG+JIMMY event generators. The latter codes account for parton showering, multiple parton interactions, beam remnants, hadronization and intrinsic p_T . The MLM algorithm is used to avoid double counting, and CKKW-style α_s -reweighting is used to smoothen the transition between the matrix-element and parton-shower phase-space parts.

3.6.6 MC@NLO

MC@NLO is a FORTRAN-based code which merges NLO matrix-element matrix-element predictions with a parton shower. The HERWIG generator the parton-shower algorithm, the hadronization model, and JIMMY can be used to simulate MPI.

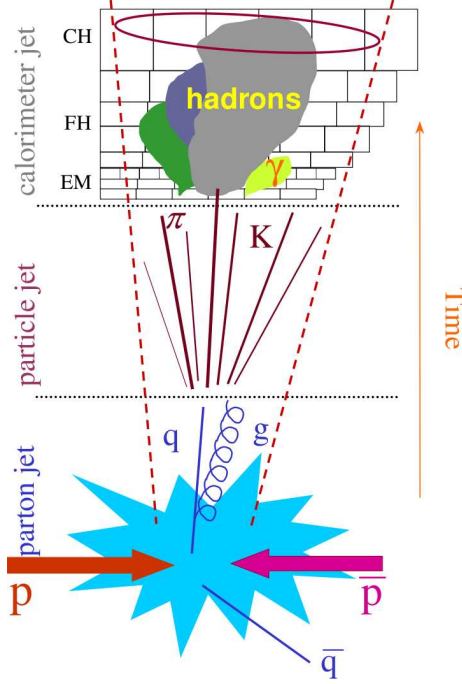


Figure 3.4: An illustration of the creation of a jet in a $p\bar{p}$ scattering, and its interaction with the experimental detector setup. The jet object can be defined as a spray of almost collinear partons, particles or energy deposits in the calorimeter. (From Ref. [58].)

3.7 Connecting Partons, Particles and Energy-Deposits in the Detector

Due to the confining property of the strong interaction the final-state objects of a parton-level calculation are not observable experimentally. Nevertheless, the topology of the observable particles arising from QCD interactions contain structures which can be identified with partons, and these structures are defined using a so-called *jet algorithm*. A jet algorithm uses a list of four-momentum vectors as input. Four-momentum vectors which are *close* to each other, defined by some metric, are grouped together to form *jets*. For a simulated event generated using an event generator code, jets defined from the list of stable particles in an event can be identified with jets formed from the list of partons that were given as input objects to the hadronization model. In this way, observable particle-level jets are mapped onto non-observable parton-level jets. The properties of parton-level jets are computable using only perturbation theory and PDFs, thereby allowing for comparison between experimental measurements of QCD related properties and the predictions of the Standard Model Lagrangian with a minimal dependence on phenomenological models like hadronization. Jet algorithms are also used to define jets from clusters of energy-deposits in so-called calorimeter sub-detector systems used in high-energy-physics experiments for amongst others for detecting jets (see Chapters 4 and 5). An illustration of jets at the parton, particle and detector levels is given in Fig. 3.4.

The main property which characterizes a jet algorithm is the metric which defines the distance between pairs of input objects. Another property of the algorithm is how the four-momentum vector of a jet is defined in terms of the four-momentum vectors of the input objects (*recombination scheme*). The most common recombination is that the jet four-momentum vector is given as the sum of the four-momentum vectors of the input objects which belong to it. In case a single input object can be clustered into two different jets, the jet algorithm must specify how to distribute the particle four-momentum vector between the two jets.

For a jet algorithm to define a mapping between parton-level, particle-level and detector-level jets which is as close as possible to one-to-one, both in terms of the number of jets and of their four-momentum vectors, the algorithm should have the property of being *collinear and infrared safe*. Infrared safe means that the list of identified jets should not change if a low-energetic object is added to the list of input objects. Collinear safe means that the list of identified jets does not change if one input object is replaced by two collinear objects with the same four-momentum vector sum as the original particle. For a jet algorithm to be usable its implementation as computer code must run fast enough, i.e. be CPU efficient enough, to allow for the analysis of a large number of events in a reasonable amount of time.

3.7.1 Cone Jet Algorithms

A cone-jet algorithm starts out with a list of initial directions (*seeds*). A geometrical cone with starting-point at the collision point and with a opening half-angle R is pointed in each of the initial directions (*trial cone*). All input objects which satisfy $\Delta R(\text{cone axis, input object}) < R$ are said to *belong* to the cone. The most commonly used recombination scheme defines the four-momentum vector of the cone as the sum of the four-momentum vector of all input objects which belong to it. If the four-momentum of the cone is parallel to the cone axis, the cone is said to be *stable*. If the cone is not stable, the axis specified by the four-momentum of the cone defines the direction of a new trial cone, and the process is repeated until a stable cone is found. This procedure is performed for each seed, resulting in a list of stable cones. Nothing prevents an input object from laying inside of two different stable cones, and modern cone-jet algorithms rely on the principle of *split-and-merge* to specify how to treat this situation: if the ratio of the shared p_T to the p_T of the softest jet is larger than a parameter f , the two are merged – if not each overlap-particle is associated to the closest jet. Due to merging, a jet can contain particles for which $\Delta R(\text{particle, jet-axis})$ exceeds R . The parameter f is referred to as the *split-and-merge* fraction.

In the simplest cone-jet algorithms the list of initial directions is taken to be the list of directions of the input objects. Such an algorithm is not infrared safe as demonstrated by a hypothetical event consisting of two high- p_T objects with a separation of $R < \Delta R < 2R$. Each object specifies a trial cone, but each trial cone only contains one object. If a low- p_T object is added inbetween the two original objects the corresponding trial cone will contain both high- p_T objects, leading to the identification of one jet instead of two. In the following, two cone-jet algorithms which address this issue are presented.

DØ Run II Mid-Point Cone-Jet Algorithm

The DØ Run II Mid-Point cone-jet algorithm [39] introduces a preclustering stage to reduce the needed CPU time per event. The list of input objects, e.g. partons, particles or calorimeter objects, is sorted in p_T , and all objects inside a $R = 0.3$ cone around the highest- p_T object

are removed from the list of input objects and are combined to form the first precluster object. The process is repeated until the list of input objects is empty. All precluster objects are required to fulfill $p_T > 1$ GeV. The list of initial directions is given by the direction of all precluster objects, as well as all directions pointing to the mid-point between each pair of precluster objects. The addition of these mid-point directions to the list of initial direction improves the infrared stability of the algorithm. The recombination scheme is given by four-momentum vector addition, and split-and-merge is performed using $f = 0.5$.

SISCone Jet Algorithm

The SISCone jet algorithm [59] is a cone-jet algorithm using the vertices of a two-dimensional (y, ϕ) grid as initial directions, with y being the rapidity (y) and ϕ the azimuthal angle. This assures a fully infrared safe algorithm since the list of trial cones does not depend on the list of input objects. Previous cone-jet algorithms which were independent of the list of input objects failed to be useful due to a prohibitively large consumption of computer power. The SISCone algorithm by-passes this problem by using a geometrically motivated clustering method which constraints the amount of computer power needed per event. In order to resolve situations where one input object belongs to two different jets, the same split-and-merge algorithm is used as in the DØ Run II Mid-Point algorithm.

3.7.2 Other Jet Algorithms

The k_T jet algorithm [60, 61] uses a different metric than cone-jet algorithms. Instead of relying on only the angle between two objects, it defines the distance between two particles approximately as the transverse momentum of the lowest- p_T particle with respect to the highest- p_T particle. This metric is related to the splitting kernels used in parton-shower algorithms: the probability of a given parton-shower splitting to be generated is inversely proportional to the k_T distance between them. This property makes the k_T algorithm suitable for analytical resummation calculations, and it is also why this algorithm is used in the CKKW algorithm for reconstructing a parton-shower history for final-state particles generated according to a matrix-element calculation (see Sect. 3.3.2). The measure of distance in the k_T algorithm between a particle i and another particle j is

$$d_{ij} = \min(p_{T,i}^2, p_{T,j}^2) \frac{\Delta R_{ij}^2}{D^2} \quad (3.6)$$

where $p_{T,i}$ and $p_{T,j}$ are the transverse momenta of the two particles, ΔR_{ij} the opening angle between the two, defined in terms of ϕ and y , and D is a unit-less parameter of the algorithm. In addition, the distance between the particle i and the beam line is defined as

$$d_{iB} = p_{T,i}^2. \quad (3.7)$$

The algorithm starts with a list of *current* particles being equal to the list of all input particles. The steps of algorithm are as follows:

1. For each i in the list of current particles, compute d_{iB} and d_{ij} for all possible choices of j , and sort the resulting d -values.
2. If the smallest d -value is a d_{ij} value, the two particles are replaced in the list of current particles by their four-momentum vector sum.

3. If the smallest d -value is a d_{iB} value, the particle i is a jet on its own and is removed from the list of current particles and added to the list of identified jets.
4. Repeat from step 1 until the list of current particles is empty.

The parameter D determines the size of the resulting jets and is analogous to the cone-size R of cone-jet algorithms: if no particle j with $\Delta R_{i,j} < D$, the distance d_{iB} for a particle i is smaller than all values d_{ij} , and the particle i forms a jet on its own according to step 3 in the algorithm above. A feature of the k_T algorithm is that there is no lower limit on the transverse momentum of the identified jets. Therefore, an additional parameter p_T^{cut} , specifying a lower transverse momentum criterion that jets have to satisfy is often introduced.

The Aachen/Cambridge jet algorithm [62, 63] is analogous to the k_T algorithm, but with the metric being

$$d_{ij} = \frac{\Delta R_{ij}}{D}. \quad (3.8)$$

As for the k_T algorithm, the parameter D defines the size of the identified jets. The pair of particles i and j with the lowest value of d_{ij} are replaced by their four-momentum vector sum. If, for a particle i , no j exist for which $d_{ij} < 1.0$, i.e. for which $\Delta R_{ij} < D$, i is a jet. The process is repeated until the list of current particles is empty, and as for the k_T algorithm a minimum jet p_T can be required. Whereas the clustering of the k_T algorithm is ordered in relative transverse momentum, the Aachen/Cambridge algorithm gives an angular ordered clustering sequence.

Final-state objects	Process	Perturbative order	α_s^0 terms	α_s^1 terms	α_s^2 terms	α_s^3 terms
Partons	Z/γ^* inclusive	LO	x			
		NLO	x	x		
		LO+LL	x	LL	LL	LL
		NLO+NLL	x	x	LL+NLL	LL+NLL
	Z/γ^* +parton	LO		x		
		NLO		x	x	
	Z/γ^* +2 partons	LO			x	
		NLO			x	x
	Z/γ^* +3 partons	LO				x
Particles	Parton-shower (PYTHIA, HERWIG)		x	LL	LL	LL
	Parton-shower with matrix-element matching (PYTHIA, HERWIG)		x	real / LL	LL	LL
	NLO + parton shower (MC@NLO, POWHEG)		x	x	LL	LL
	CKKW / MLM (SHERPA, ALPGEN+PYTHIA)		x	real / LL	real / LL	real / LL

Table 3.1: An attempt to illustrate which parts of the full, all-order perturbative interaction amplitude are taken into account in the various models described in the main text, for the example process of $Z/\gamma^*(+jets)$ production. The label x indicates that all terms to the given order which are relevant for the process in question are taken into account. The label LL ($LL+NLL$) indicates that only the leading (leading and next-to-leading) logarithmically enhanced terms are included at the given order. The label *real* means that only real-emission terms are taken into account at the given order, i.e. that virtual-correction terms arising from loop diagrams are ignored.

Chapter 4

Experimental Setup: Tevatron and the DØ Detector

The Tevatron accelerator [64], located at the Fermi National Accelerator Laboratory [65] nearby Chicago, USA, delivers proton-antiproton collisions for two multipurpose experiments, CDF and DØ. From 1992 to 1996 the Tevatron operated at a centre-of-mass energy of 1.8 TeV (Run I). Each experiment recorded data sets corresponding to an integrated luminosity of 120 pb^{-1} , leading to the discovery of the top quark [66, 67] in 1995. After a major upgrade of both the accelerator and the detectors, a second period of data taking was started in 2002 (Run II). Until today each experiment has collected data sets corresponding to an integrated luminosity of $\sim 6 \text{ fb}^{-1}$. Until the end of Run II, which is foreseen to last until 2010 or 2011, this is expected to increase to $8\text{-}10 \text{ fb}^{-1}$.

4.1 The Tevatron Accelerator

The Tevatron is the last step in a long chain of accelerators, see Fig. 4.1. A Magnetron transforms hydrogen gas into H^- ions which are accelerated first to 0.75 MeV by a Cockcroft-Walton accelerator and then to 400 MeV by a linear accelerator referred to as the *LINEAC*. The H^- ions are passed through a thin carbon foil in order to remove the electrons, leaving a beam of protons. The protons are accelerated to 8 GeV in a circular synchrotron (circumference 0.2 km) referred to as the *Booster*. The next and second-last step of the proton acceleration chain is the *Main Injector*, which is a circular synchrotron with a circumference of 3.3 km. The Main Injector delivers 150 GeV proton beams to the Tevatron accelerator, and is also the initiator of the antiproton generation chain. A 120 GeV proton beam is collided with a fixed target made of nickle, and antiprotons with an energy of $\sim 8 \text{ GeV}$ are extracted from the arising spray of hadrons using a charge-momentum spectrometer. The antiprotons are turned into a manageable beam using the *Debuncher* synchrotron, removing the bunch structure inherited from the initiating proton beam while providing a uniform energy of 8 GeV. Both the Debuncher and the following *Accumulator* storage synchrotron utilize stochastic cooling [68] to reduce the spread in transverse and longitudinal energy of the antiprotons. The Accumulator is followed by the *Recycler* which shares tunnel with the Main Injector. The purpose of the Recycler is to recycle antiprotons from the Tevatron, and to enrich the recycled beam with “new” antiprotons from the Accumulator. Due to technical problems early in Run II, the Accumulator was long the sole provider of antiprotons to the Recycler, but since 2005 the Recycler also accepts antiprotons from the Tevatron remaining from a previous fill. The Accumulator has a large storage capacity for 8 GeV antiprotons and applies both stochastic cooling and electron cooling [69] to further increase the uniformity of the beam. Once a sufficiently large number of antiprotons has been accumulated in the Recycler the antiprotons are transferred to the Main Injector and accelerated to 150 GeV. The

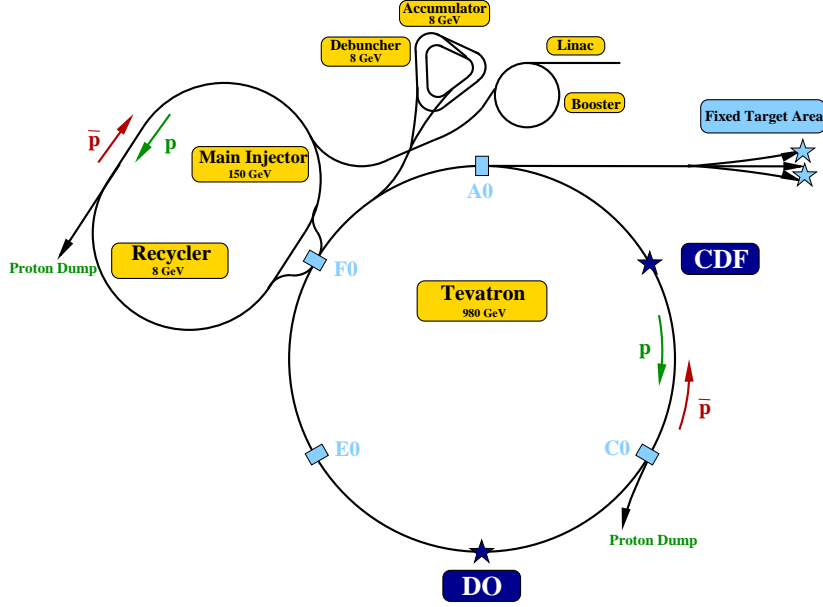


Figure 4.1: Graphical representation of the $p\bar{p}$ accelerator chain which culminates in the Tevatron.

Main Injector delivers beams of protons and antiprotons to the main Tevatron accelerator (circumference 6.28 km), which increases the energy to 980 GeV and provides collisions for the CDF and DØ experiments. The Tevatron beams circulate in opposite directions with an orbit time of $20.9 \mu\text{s}$. Each beam is split into 36 bunches of protons/antiprotons. A bunch contains of the order of $2.5 \cdot 10^{11}$ ($5.4 \cdot 10^{10}$) protons (antiprotons), and has a length of ~ 2 ns or 60 cm. The bunches are not uniformly distributed around the ring, but are located in 3 so-called super bunches, each separated by $2.6 \mu\text{s}$. A super bunch contains 12 bunches, each separated by 396 ns. Due to interactions of a beam with the other beam, with beam gas or with the beam pipe, the quality of the beams reduces with time. The limiting factor for the Tevatron luminosity is the availability of antiprotons. Accumulating enough antiprotons in the Recycler to inject a new beam takes about 20 hours, and this defines how long each store lasts.

4.2 The DØ Detector

The DØ detector is a multipurpose detector designed to measure the kinematic properties of the final state particles in $p\bar{p}$ collisions delivered by the Tevatron accelerator, and to be able to distinguish between electrons, muons, taus, photons and hadron jets. The detector provides nearly 4π solid angle coverage and follows the standard design with cylindrical layers of subdetector, see Fig. 4.2. The first layer – the central tracking system – detects the trajectories of charged particles and contains a magnetic field which enables the measurement of the momentum of a track from its curvature. The second layer – the electromagnetic (EM) calorimeter – measures the energy of electrons and photons, as well as parts of the energy of hadron jets. The third layer – the hadronic (HAD) calorimeter – measure the remaining energy of hadron jets. Only muons and neutrinos reach the outermost, forth layer – the muon

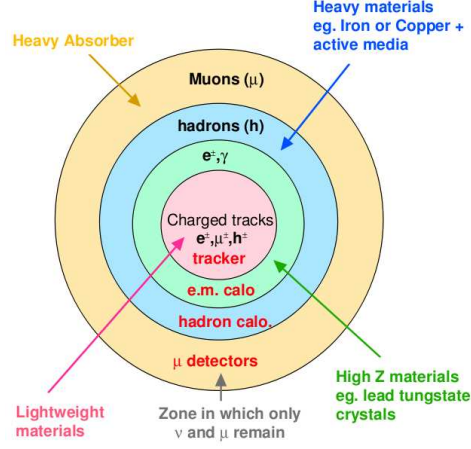


Figure 4.2: Sketch of a generic multipurpose high-energy physics detector [70].

system – which is used to detect muons. The muon system contains a magnetic field, which, in combination with the central tracking system, allows for a measurement of muon momenta. The transverse components of the sum of energies carried by neutrinos is indirectly measured through momentum conservation as the opposite of the vector sum of all energies directly measured in the event, thus completing the measurement of all final state particles predicted by the Standard Model.

A cross-sectional overview of the DØ Run II detector is given in Fig. 4.3. The following sections contain a brief overview of the DØ detector based on [71], concentrating on the aspects relevant for the analysis presented later in this thesis. In the description of the detector, and later in the analysis, a right-handed coordinate system is used where the z -axis points in the direction of flight of the proton beam, and the y -axis points upwards (see Fig. 4.3).

When referring to the position of detector components (*detector coordinate system*), the origin is taken to be the geometrical center of the detector. The corresponding polar coordinate system is given by the distance to the origin and the azimuthal angle, ϕ_{det} , and the polar angle, θ_{det} . Due to the cylinder symmetry of the detector, r , the distance from the z axis is often used instead of the distance to the origin. The pseudo-rapidity, defined by $\eta_{\text{det}} = -\ln[\tan(\theta_{\text{det}}/2)]$, approximates the true rapidity $y = \frac{1}{2} \ln[(E + p_z c)/(E - p_z c)]$, for finite angles in the massless limit ($\frac{m}{E} \rightarrow 0$). When referring to the kinematic properties of the $p\bar{p}$ collisions final state, the so-called *physics coordinate system*, where origin is the primary $p\bar{p}$ interaction vertex, is used. The corresponding angles are labelled, ϕ_{phys} , θ_{phys} and η_{phys} , or just ϕ , θ and η .

4.2.1 Central Tracking System

The aim of the central tracking system is to accurately measure the position of charged particles as they cross the various detector layers (*hits*), with a minimal energy loss for the particles. The hits can be used to reconstruct the trajectories of charged particles. The central tracking system is located inside a 2 T solenoidal magnetic field allowing track momenta to be deduced from the curvatures of the tracks. Tracking information is used for

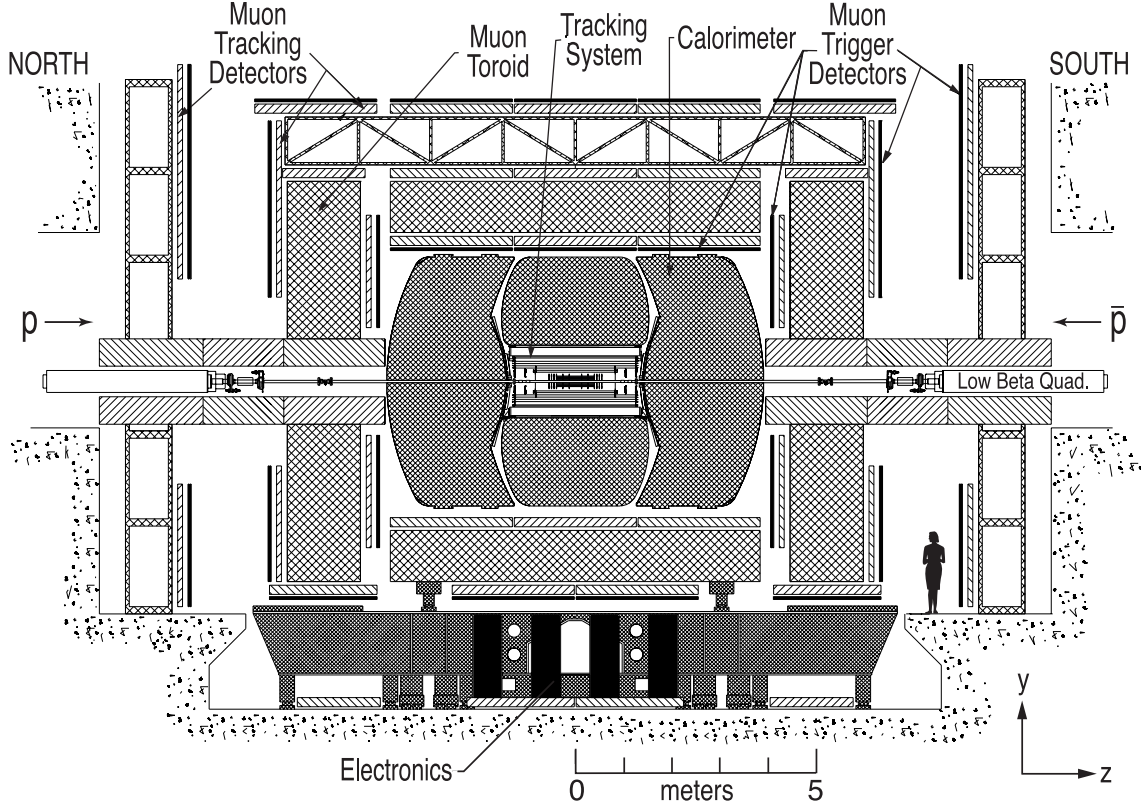


Figure 4.3: The DØ Run II detector [71].

the identification of the primary interaction vertex, electrons and muons, as well b quarks via the decay vertices of relatively long-lived b hadrons. An overview of the central tracking system is given in Fig. 4.4.

SMT

The *silicon microstrip tracker* (SMT) is the first detector system encountered by particles after they enter the DØ detector. Its purpose is to provide accurate tracking and vertexing for $\eta_{\text{det}} < 2.5$. The relatively broad interaction vertex distribution, whose width is $\sigma_z \approx 25$ cm, guides the choice of geometric layout for the SMT. The silicon modules contain about 10 cm long strips which give a signal if crossed by a charged particle. Neighbouring layers have different angular orientation of the strips, allowing for the precise measurement of $(\eta_{\text{det}}, \phi_{\text{det}}, r)$ points along the trajectories of charged particles. The silicon modules are mounted on *barrel* and *disks*, see Fig. 4.5. Six four-layered barrels with a length of 12 cm and $2.7 < r < 11$ cm are capped at high $|z|$ by so-called *F*-disks. Forward of the three disk+barrel pairs, units of three more discs follow at $43 < |z| < 53$ cm. The purpose of the barrels is to provide accurate tracking at low values of η_{det} . The disks fulfill this task for high η_{det} , together with additional pairs of two so-called *H*-disks at $|z| = 100$ cm and 121 cm. The latter disks act as a high- η_{det} extension of the *central fiber tracker* (CFT, see below).

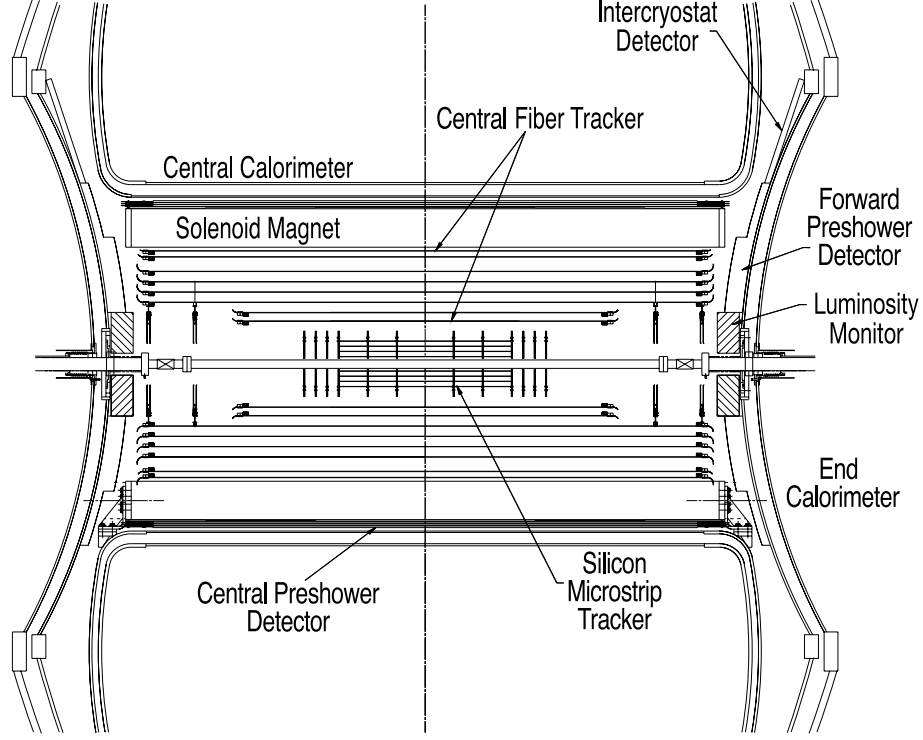


Figure 4.4: Cross-sectional overview of the central tracking system [71].

CFT

The CFT provides tracking using long, scintillating fibers mounted on eight cylinders at $20 < r < 52$ cm. The two innermost cylinders have a length of 1.66 m to make room for the outermost SMT disks, whereas the six outer cylinders are 2.52 m long. The CFT provides tracking up to $|\eta_{\text{det}}| \approx 1.7$. Each cylinder supports doublet layers of $835 \mu\text{m}$ thick fibres oriented along the beam axis, and a second doublet layer shifted by $\pm 3^\circ$ in ϕ_{det} , with the sign alternating between the different layers. Clear readout fibres are attached to one end of the scintillating fibres; the other ends are mirrored. The readout fibres transfer the light signal to photomultipliers which have a quantum efficiency of $\geq 75\%$ and are capable of detecting single photons. In total, the CFT uses about 200 km of scintillating fibres and 800 km of clear readout fibres.

Solenoidal Magnet

The central tracking system contains a homogeneous magnetic field of 2 T provided by a superconducting solenoidal magnet made of Cu:NbTi which is situated at $r \approx 60$ cm and has a length of 2.7 m. The solenoidal system was designed to optimize the tracking momentum resolution and pattern recognition capabilities while at the same time limiting the amount of material in front of the calorimeter. The magnet, including the cryostat, has a thickness of about one electromagnetic radiation length (X_0) at $\eta = 0$. The superconductors are operated with a current of 4.7 kA at a temperature of ~ 4 K using liquid Helium cooling, well below the critical temperature of NbTi (~ 10 K). The time needed to cool down the 1.5 tonne structure

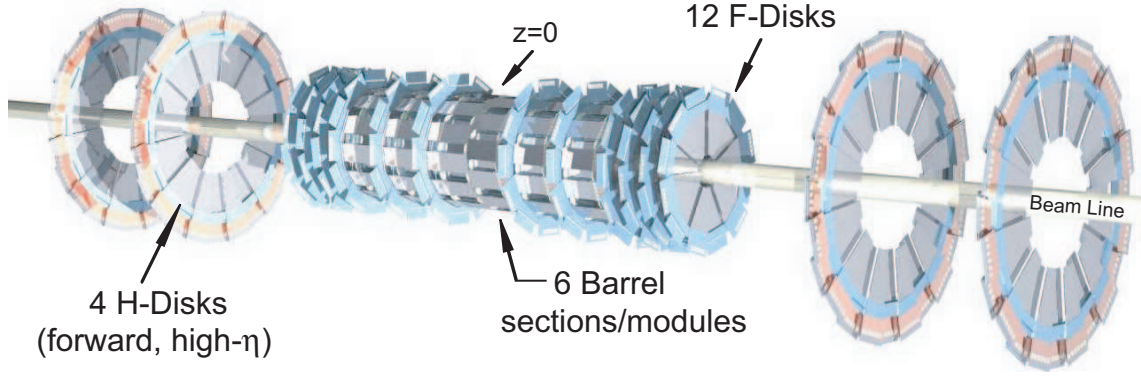


Figure 4.5: The geometric layout of the SMT system, consisting of barrels and disks carrying silicon sensors [71].

from room temperature to operating temperature is about 40 hours, and the process leads to a contraction of the physical dimensions of the structure by 0.4% [72].

4.2.2 Calorimeter System

Having measured the trajectories of charged particles, the calorimeter aims to fully absorb and measure the energy of all particles, except for muons and neutrinos. The calorimeter also aids in separating hadronic jets from electrons and photons. The overall transverse energy balance in each event is also measured using the calorimeter, with help of the muon systems (see below), thus providing an indirect measurement of the transverse momentum sum of the neutrinos in the event and any non-SM particles escaping detection.

The calorimeter is divided into three section, each housed in its own cryostat: the central calorimeter (CC) covering $|\eta_{\text{det}}| \lesssim 1.0$ and two endcap calorimeters (EC) extending the coverage up to $|\eta_{\text{det}}| \approx 4.0$. In each section, the four innermost layers make up the electromagnetic (EM) calorimeter. Further out are the fine hadronic (FH) and coarse hadronic (CH) calorimeters. The calorimeters are composed of three main elements: passive absorber plates which initiates particle showers; liquid Argon which acts as the active material, sampling the shower via ionization; and copper readout plates on which the ionization electrons are collected using a 2 kV electric fields. The absorber plates in the EM (FH; CH) calorimeter are made of 3 – 4 mm (6 mm; 47 mm) thick layers of depleted uranium (uranium-nobbiium; copper (CC) and stainless steel (EC)). Several cells with similar $(\eta_{\text{det}}, \phi_{\text{det}})$ coordinates share the same readout channel, forming a readout cell. Similarly, several readout cells are combined into projective calorimeter towers, see Fig. 4.6. The boundary regions between each tower, referred to as ϕ -cracks, have reduced energy sensitivity. The calorimeter has a granularity of $\Delta\eta_{\text{det}} \times \Delta\phi_{\text{det}} = 0.1 \times 0.1$. An exception is the third EM layer, where the peak of the EM showers are expected, which has a finer resolution of 0.05×0.05 to improve the directional calorimeter measurement of electrons for matching with tracks. Electromagnetically interacting particles have a mean free path of flight in uranium of $X_0 \approx 3.2$ mm. The EM calorimeter has a depth of $\sim 20X_0$, meaning that most EM showers deposit their entire energy without reaching the hadronic calorimeter. Hadronic particles have a significantly weaker interaction with uranium, and as a result the EM calorimeter accounts for only $0.8\lambda_A$, with λ_A being

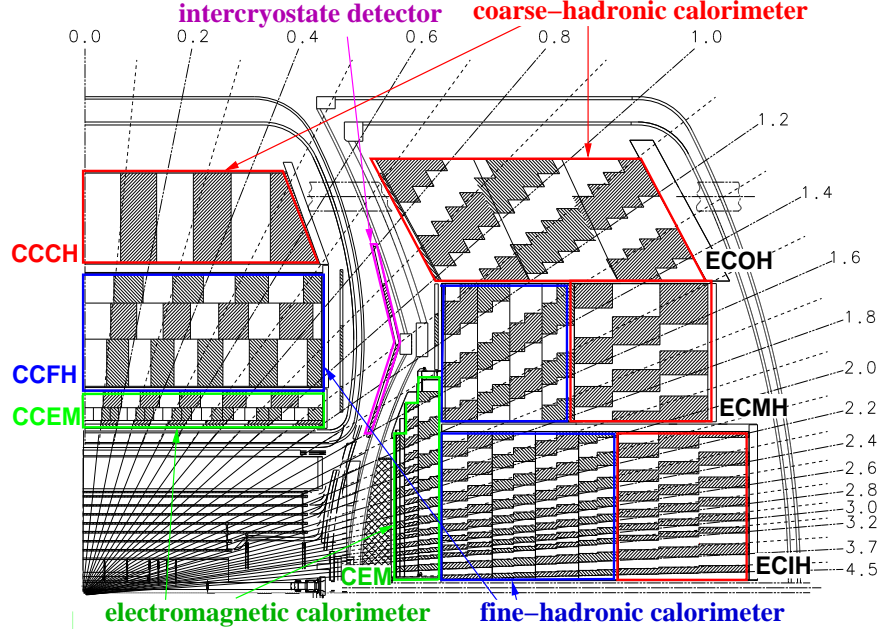


Figure 4.6: The structure of the DØ calorimeter [71].

the nuclear interaction length (or the mean free path of hadronic particles with respect to gluon bremsstrahlung). The total depth of the calorimeter corresponds to $7\lambda_A$, containing more than 99% of all hadronic energy in the event. Hadronic particles interact with the absorber plates, generating mostly charged and neutral pions. Only neutral pions, decaying to two photons, gives a significant amount of ionization in the liquid argon. One consequence of this is that statistical fluctuations in the amount of π^0 s created in hadronic showers lead to an intrinsically worse energy resolution for jets than for electrons or photons. Another consequence is that hadronic energy showers have a broader lateral extension than pure EM showers, a fact that is exploited to distinguish hadronic jets from electrons and photons.

4.2.3 Muons System

The three-layered muon system uses proportional drift tubes to provide accurate tracking for $|\eta| \lesssim 1.0$, with fast scintillators being used for triggering and for timing information to associate muons with a specific bunch crossing. The coverage is extended to $|\eta| \lesssim 2.0$ using mini drift tubes in addition to scintillators. A toroid, located between the first and the two last muon system layers, provides a 1.6 – 1.8 T magnetic field to allow for momentum measurements independent of the central tracker. The muon system lacks coverage in the downward directions where the support structures of the DØ detector are located ($4.25 < \phi_{\text{det}} < 5.15$).

4.2.4 Luminosity Counter

The main task of the luminosity counters is to determine the instantaneous Tevatron luminosity at the DØ interaction point by measuring the total inelastic $p\bar{p}$ cross section. This is done

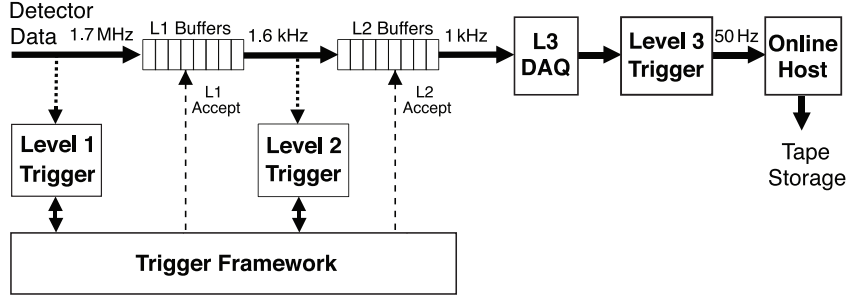


Figure 4.7: Schematic overview of the DØ trigger and data acquisition system [71].

using two arrays of 24 plastic scintillator counters at $z = \pm 140$ cm covering $2.7 < |\eta_{\text{det}}| < 4.4$ (see Fig. 4.4). Additional tasks of the luminosity counters are to provide a fast measurement of the z coordinate of the interaction vertex, and to measure the vertex multiplicity as well as the beam halo rates.

4.2.5 Trigger and Data Acquisition System

The Tevatron delivers $p\bar{p}$ bunch crossings with a frequency of 1.7 MHz but the event storage capacity to tape is only 50 Hz. The task of the trigger system is therefore to select the 0.03 per-mill most interesting bunch crossings which can be stored for offline analysis. The selection is done in three levels, as illustrated in Fig. 4.7.

L1

The L1 trigger is implemented using custom-designed hardware. Detector information corresponding to ten bunch crossings can be kept on hold in a buffer awaiting the L1 trigger decision, giving a decision time of $\sim 4 \mu\text{s}$ and a reduction in rate from 1.7 MHz to 2 kHz. The L1 trigger decision is based mainly on information from the calorimeter, the CFT and the muon system. Electron triggers require $\Delta\phi_{\text{det}} \times \Delta\eta_{\text{det}} = 0.2 \times 0.2$ EM calorimeter towers with E_T above a certain threshold, typically 5 – 10 GeV. Other triggers use tracking information from the CFT and from the muon system.

L2

The L2 trigger, implemented using hardware engines in combination with embedded micro-processors, reduces the event rate from 2 kHz to 1 kHz. At this level, correlations between physics objects can be used. For instance, some electron triggers require a track candidate to point towards the EM calorimeter tower, or puts an upper limit on the amount of transverse energy deposited nearby the EM calorimeter tower.

L3 and the Data Acquisition System

For events passing the L2 trigger the entire detector is read out and the digitized event information is sent to a dedicated computer farm using the L3 data acquisition system,

which has a capacity of 250 MB/s. The high level, fully programmable L3 software trigger system reconstructs physics object candidates which are similar to those used in the offline event analysis. Electron candidates are identified using $\Delta R = 0.25$ calorimeter cones, with threshold on the transverse energy, the ratio of the EM energy in the cone to the EM+HAD cone energy, as well as the electron isolation. Track candidates are reconstructed using detailed algorithms combining hits from the SMT and CFT systems, and some electron triggers require a track pointing towards the calorimeter electron candidate.

4.2.6 Offline Reconstruction

Events passing the L3 trigger are written to tape and made available for offline analysis. Offline reconstruction algorithms available through the DØreco package [73] associate electronics channels with detector modules, with the output being e.g. energy clusters in the calorimeter or hits in the tracking system. Detector-specific calibration constants are applied before physics objects candidates are identified (for details, see Chapt. 5). The resulting event samples are made available in several data formats, ranging from the DST format, containing the full event information, to the ROOT-based [74] Common Analysis Format (CAF) [75], ntuples providing a compact format with the information relevant for most physics analysis. The latter format was used for this analysis, using CAF samples provided by the DØ Common Samples Group [76].

4.3 Simulation of the Experimental Setup

To be able to understand the connection between the event information registered with the DØ detector and the physics governing the $p\bar{p}$ collisions, it is essential to have a detailed computer simulation of each step in the experimental setup. Particle-level event generators are used to simulate $p\bar{p}$ collision. A GEANT-based [77] simulation package called DØgstar [78] is used to describe the interaction between the outgoing particles and the various sub-systems of the DØ detector. The simulated energy deposits are converted into the format that real data has at this stage using the DØsim [79] package. It is essential to account for detector noise and inefficiencies, for the possibility of more than one $p\bar{p}$ scattering during the same bunch crossing, and for electrical cross talk with neighbouring bunch crossings due some detector elements, like the calorimeter, which have readout time being longer than the time delay between two consecutive crossings. These effects are accounted for by overlaying the detector information of the simulated main $p\bar{p}$ collision with the corresponding information from random bunch crossings registered with the real DØ detector using a dedicated zero bias trigger. The zero-bias overlaid simulated event samples undergo the same offline reconstruction process as used for real data.

Chapter 5

Event Reconstruction and Object Identification

The event reconstruction and object identification process translates the digitized event information from the detector into physics objects corresponding to the final-state particles of the interactions taking place during a triggered bunch crossing. It gives information about the particle types as measurements of the particle energies and directions of flight. The detector design is optimized to allow for the identification of electrons, muons, taus, photons, hadronic jets and, indirectly through the total E_T balance measured in the event, neutrinos. The present analysis relies on the accurate identification of electrons, jets and the primary interaction vertex, using information from the central tracking system and the calorimeter.

5.1 Tracks

The SMT and CFT (see Sect. 4.2.1) provide information about the position of charged particles as they passed the various layers of the tracking system. The hits in the tracker correspond to points along the trajectory of flight of the particle, and the accurate reconstruction of this trajectory is the aim of the tracking algorithm. Due to the solenoidal magnetic field in the tracking volume, the trajectory of a charged particle in the transverse plane (x, y) corresponds to a circle. In the transverse plane a track can be parametrized in terms of the curvature, ρ , the distance of closest approach (DCA) to the beam line, d_0 , and the azimuthal angle of the track direction at the point of closest approach to the beam line, ϕ . Two additional parameters are needed to describe the track in the three-dimensional space, e.g. the z coordinate and the polar angle of the track at the point of closest approach to the beam line. The DØ track reconstruction is divided into two steps. First, pattern recognition techniques are used to identify track candidates from the hits. Second, the hits of the track candidates are used as input to a fitting algorithm which extracts the track parameters.

Two different algorithms are used for the initial pattern recognition step. The Histogramming Track Finding (HTF) algorithm assumes that all relevant tracks pass, or nearly pass, through the beam line, i.e. that $d_0 \approx 0$. With this assumption a single pair of hits in the tracking system uniquely defines a track in the transverse plane, characterized by (ρ, ϕ) . A 2-dimensional histogram is filled with the (ρ, ϕ) coordinates of all pairs of tracker hits. Pairs of hits lying along the trajectory of the same charged particle share the same (ρ, ϕ) point, meaning that peaks in the histogram correspond to plausible track candidates. The HTF algorithm is based on this basic idea, but contains several modifications to assure a fast algorithm in the presence of the 10^4 to 10^6 hits typically contained in a DØ event.

A second pattern recognition algorithm, the Alternative Algorithm (AA), is also used. Initial track candidates are identified from three SMT hits, and are kept if the track has a small DCA, a good χ^2 fit to the hits and curvature corresponding to that of a circle with

radius larger than 30 cm, implying a momentum larger than 180 MeV. Each track candidate is extrapolated outwards to the next SMT or CFT layer giving a search window where a new hit is expected. If a hit is found, the track is refitted using this hit in addition to the original ones. The lack of an expected hit is counted as a missed hit, and the track candidate is rejected if it has more than 3 missed hits. A second iteration of the algorithm, using three CFT hits as starting point, is performed to catch tracks containing two or less SMT hits. To limit the needed computer time by rejecting fake track candidates, CFT-initiated tracks are required to point towards a vertex candidate (see below) reconstructed using only the already identified SMT-initiated tracks.

To extract the track parameters, candidates from the HTF and AA algorithms are used as input for a Kalman algorithm [80] which performs detailed track fits taking into account effects like the non-uniformity of the magnetic field and the energy loss of charged particles due to interaction with detector material. The curvature determined by the fit corresponds to a measurement of the track momentum.

5.2 Vertex Reconstruction

A track points towards the point in space, the *vertex*, where the charged particle was created. Knowledge of the tracks of two or more particles arising from the same vertex allows the coordinate of the vertex to be determined, which is the task of vertexing algorithms. Each $p\bar{p}$ scattering taking place during a bunch crossing gives rise to an interaction vertex. In addition, long-lived particles such as b -hadrons also give rise to vertices, referred to as *secondary* vertices. The vertex coordinates of the $p\bar{p}$ scatterings are needed to determine the transverse components of energy deposits in the calorimeter, and to associate objects to different vertices based on the direction of tracks. In $D\bar{O}$, the vertices in each event are determined using an iterative, reweighted Kalman fitting algorithm [81] following Ref. [82]. For the first iteration, all tracks are assigned a weight of unity. For iteration number n , tracks are down-weighted for the determination of each vertex based on their DCA to the vertex in question, with the vertex coordinates being those determined in iteration number $n - 1$. The motivation for the reweighting is to reduce the impact of mis-associated and mis-measured tracks on the final vertices.

Having determined a list of vertices, the next task is to select the vertex associated with the main interaction of the bunch crossing, the *primary* vertex. The primary vertex is distinguished from vertices arising from additional pile-up $p\bar{p}$ interactions using a likelihood method exploiting the higher average track p_T from the primary vertex. For each track associated with a vertex the probability that the track was created in a pile-up interaction is evaluated based on the track p_T distribution of pile-up interactions. The track p_T distribution of pile-up interactions was extracted from a data sample registered using a so-called minimum-bias trigger, i.e. a trigger which requires only a small amount of activity in the detector. For each reconstructed vertex, the probability that the vertex arised from a pile-up interaction is defined as the product of the probabilities of the single tracks associated with the vertex. For each event, the vertex which has the lowest probability for having arised from a pile-up interaction is selected as the primary interaction vertex.

5.3 Calorimeter Zero Suppression

Due to electronic noise as well as uranium decays, calorimeter cells will register a nonzero energy in the absence of any signal. The identification of physics objects can be significantly simplified by rejecting all cells which have less energy than N times the noise-only energy RMS value (*zero suppression*). In DØ, only cells satisfying at least one of the two following criteria are used for object identification:

- $E_{\text{cell}} > 4 \cdot \text{RMS}_{\text{noise}}$, or
- $E_{\text{cell}} > 2 \cdot \text{RMS}_{\text{noise}}$, and the cell is a neighbor of a cell satisfying $E_{\text{cell}} > 4 \cdot \text{RMS}_{\text{noise}}$.

This algorithm is referred to as *T42* and is documented in Ref. [83]. In addition to reducing the complexity of the calorimeter object reconstruction process, T42 has been found to give a 6% better jet energy resolution than achieved using a simple $2.5 \cdot \text{RMS}_{\text{noise}}$ zero suppression threshold [84].

5.4 Electrons

The identification of electron candidates starts with running a simple $\Delta R = 0.2$ cone algorithm using calorimeter towers based on the four EM and the first hadronic calorimeter layers as input objects. Towers satisfying $E_T > 0.5$ GeV are used as seeds. Identified EM candidates are divided in two categories based on whether or not a track candidate is pointing towards the calorimeter tower (*track match*).

The electron identification process separates electrons from hadrons and photons using detailed information from the calorimeter and the central tracking system (see Chapt. 4). Compared with hadronic jets, electrons are characterized by a narrower lateral shape of the calorimeter shower. The shower shape of electrons is nearly identical with that of photons, but the two can be distinguished using tracking information: an electron calorimeter object normally has one and only one track match, whereas a photon has zero, or, in the case of a $\gamma \rightarrow e^+e^-$ conversion, two track matches. A single π^0 particle, with $\pi^0 \rightarrow \gamma\gamma$, has a similar calorimeter deposit as electrons and photons, and if a π^0 overlaps in direction of flight with a charged hadron, this can result in a fake electron candidate. However, in contrast with a real electron, the fake candidate usually has a track momentum which differs significantly from the calorimeter energy, corresponding to $E_T^{\text{calo}}/p_T^{\text{track}}$ deviating from unity. In case $E_T^{\text{calo}}/p_T^{\text{track}}$ is by chance close to unity the charged hadron will normally lead to a significant fraction of the electron candidate energy being reconstructed in the hadronic calorimeter, in contrast to what is the case for real electrons.

5.4.1 Electron Identification

The electron preselection requires an EM candidate satisfying the isolation criterion

$$iso = \frac{E_{EM+HAD}^{R=0.4} - E_{EM}^{R=0.2}}{E_{EM}^{R=0.2}} < 0.2, \quad (5.1)$$

where $E_{EM+HAD}^{R=0.4}$ ($E_{EM}^{R=0.2}$) is the energy in the entire calorimeter (EM calorimeter) in a cone of $R = 0.4$ ($R = 0.2$) around the center of the electron candidate. Additionally, the

candidate's electromagnetic fraction,

$$EMfrac = E_{EM}^{R=0.2} / E_{EM+HAD}^{R=0.2}, \quad (5.2)$$

must be larger than 0.9. The main electron selection uses a likelihood discriminant [85] utilizing seven calorimeter and tracking variables. The likelihood is only defined for electron candidates which have a track match. The input variables are:

1. distance of closest approach to the primary vertex, defined in the transverse plane;
2. number of tracks with $p_T > 0.5$ GeV in a cone of $R = 0.05$ around the electron track;
3. p_T sum of all tracks in a cone of $R = 0.4$ around the electron track;
4. electromagnetic energy fraction, $EMfrac$;
5. χ^2 -probability for the match between the track and the calorimeter cluster;
6. ratio E_T^{calo} / p_T^{track} ;
7. the HMxmatrix, $Hmx7$,

where $Hmx7$ is a χ^2 discriminant used to suppress the contribution from fake candidates arising from hadrons. It depends on seven calorimeter variables: the total EM energy, the vertex z position, the shower width in ϕ and the energy fraction contained in each of the four EM layers of the calorimeter. The distributions of four of the likelihood input variables for signal and background samples are given in Fig. 5.1. The resulting distributions of the likelihood discriminant for electron candidates in the central and endcap calorimeters are given in Fig. 5.2.

5.4.2 Electron Identification Efficiency

The analysis presented in this thesis relies on the identification (ID) efficiency being identical for simulation and data. That this is the case can be verified by measuring the ID efficiency separately in data and simulation and then correcting the simulation for any observed discrepancies. To measure the electron ID efficiency one must

1. know the direction of flight of a final-state electron, and
2. check if an electron candidate was identified in the corresponding part of the detector.

A way of achieving 1) is the *tag-and-probe* method. This requires the identification of a data sample where the presence and kinematics of an object can be established using only information which is uncorrelated with the selection criteria whose efficiency is to be measured. For electrons, this is done using a $Z/\gamma^* \rightarrow ee$ sample. One electron, the *tag*, is identified using tight ID criteria to reduce the amount of background. The tag electron is required to have fired a trigger to avoid any potential trigger bias. The presence and kinematics of the second electron, the *probe*, is established by requiring a track candidate which satisfies that

- the track points to the primary vertex,
- the track is isolated from all other tracks,

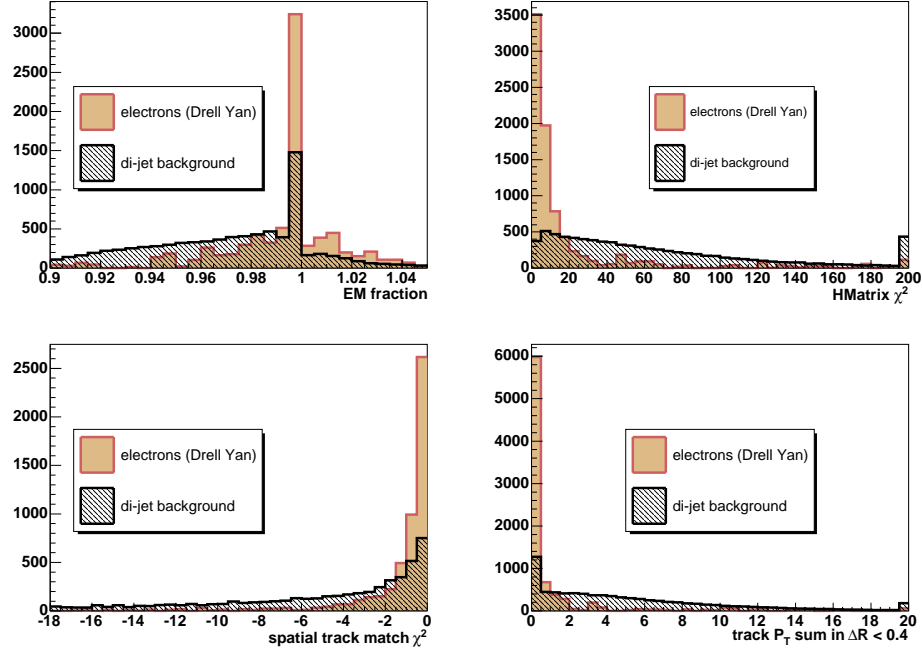


Figure 5.1: Distributions of four electron likelihood input variables for signal (shaded) and background (dashed) [86]. The background sample consists of dijet events where a jet is reconstructed as an electron candidate. Electron candidates can have $EMfrac > 1.0$ due to the possibility of a negative hadronic energy arising from noise.

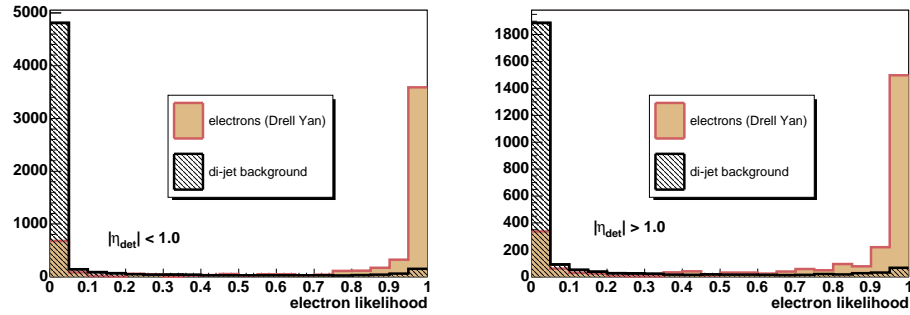


Figure 5.2: Distributions of the electron likelihood discriminant for electrons (shaded) and jets (dashed) using the same samples as for Fig. 5.1 for $|\eta_{det}| < 1.1$ (central calorimeter) and $1.5 < |\eta_{det}| < 2.5$ (endcap calorimeter) [86].

- the track is back-to-back in ϕ with the tag electron, and
- the invariant mass of the tag and the track is $\approx M_Z$.

It is then tested if

- a) an electron candidate is found in the direction of the track,

and if

- b) this candidate satisfies the selection criteria whose efficiency is to be determined.

The probability that a) is satisfied defines the *preselection efficiency*, or the *reconstruction efficiency*. The conditional probability that b) is satisfied, given a), is referred to as the *post-preselection ID efficiency*, or just the *ID efficiency*. Each event contains two electrons, and to avoid a bias both electrons in every event are treated as a potential tag electron. The tag-and-probe method is designed to measure the electron ID efficiency using only detector information. However, the preselection of the tag-and-probe measurement might bias the measurement. The size of this bias was evaluated using a simulated event sample. The efficiency measured in the simulated sample using the tag-and-probe method was compared with the efficiency defined using generated electrons as probe, i.e. the true efficiency, and the difference was found to be negligible [87].

In the present analysis reconstructed electrons are defined by requiring the standard preselection in addition to an electron likelihood value larger than 0.2 and $p_T > 15$ GeV. The corresponding preselection and ID efficiencies for data and simulation are available from the DØ EM ID group [88] and are shown as a function of η_{det} in Fig. 5.3.

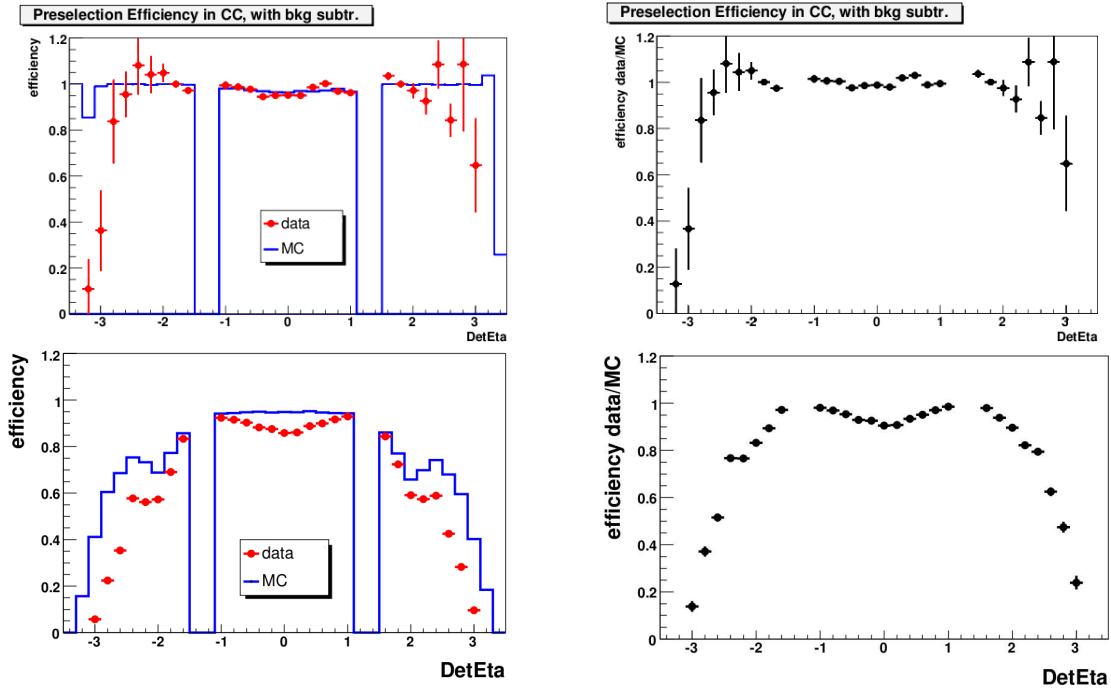


Figure 5.3: Preselection (upper left) and ID (lower left) efficiency of electron selection criteria, as a function of η_{det} , for data and simulation. The correction factors to be applied to simulation to correct for the discrepancies are given for the preselection efficiency (upper right) and ID efficiency (lower right). (Adopted from Ref. [89].)

5.4.3 Electron Energy Scale and Resolution

The task of the electron energy scale calibration is to translate the calorimeter signals into electron energies in units of GeV. The true electron energy, E^{true} , can be related to the digitalized calorimeter signal in each of the five calorimeter layers, A_i , through

$$\begin{aligned} E^{\text{measured}} &= \alpha \cdot \sum_{i=1}^5 \beta_j A_i + \delta \\ &= \alpha \cdot E^{\text{reco}} + \delta, \end{aligned} \quad (5.3)$$

where β_j is the relative response correction of the five layers with $\beta_3 \equiv 1.0$, α is the overall calibration factor, δ is an energy offset correction and $E^{\text{reco}} \equiv \sum_{i=1}^5 \beta_j A_i$. A non-zero δ can arise from energy loss in material before the calorimeter, noise, energy from additional $p\bar{p}$ interaction and from signal being lost due to zero suppression. The absence of $O(A_i^N)$ ($N > 1$) terms in Eq. (5.3) is a good approximation for electron energies above 10 GeV [90]. In Run I, an estimate of the EM scale was determined from test-beam measurements, giving a reconstructed Z boson mass of $M_Z = 87.11 \pm 0.18$ GeV [91]. To achieve a better precision, the world-average value $M_Z = 91.1876 \pm 0.0021$ GeV [1] is taken as input, with $J/\Psi \rightarrow e^+e^-$ and $\pi^0 \rightarrow \gamma\gamma$ having been used for cross-checks [90, 92]. Assuming the validity of Eq. (5.3), the EM energy scale can then be calibrated to a high level precision. The parameters α , β_j and δ can be determined such that the reconstructed dilepton mass, M_{ee} , is equal to M_Z [93]. However, this technique ignores higher-order QED corrections ($Z \rightarrow ee\gamma$) which invalidates the assumption that $M_{ll} = M_Z$ if $\Delta R(e, \gamma)$ is larger than the cone size used for defining reconstructed electron candidates, i.e. $\Delta R > 0.2$ for DØ. To include the impact of higher-order QED corrections, a full event generator is used, with M_Z^{PDG} as input. A parametrized detector description, depending on the unknown α and β as well as the energy resolution, is used to account for the impact of the detector on the M_{ee} spectrum [92, 93, 94]. Assuming a reasonable value for the energy resolution, a likelihood method is used to determine the α , β_j and δ values which give the best agreement between the simulated and the reconstructed M_{ee} distributions. In a second step α , β_j and δ are kept fixed, and the energy resolution is determined in the fit. The resulting electron energy scale has an uncertainty less than 0.5% [93].

5.5 Jets

Jet candidates are reconstructed using the DØ Run II Mid-Point Cone Jet algorithm (see Sect. 3.7.1), using calorimeter towers after zero suppression as input objects. Jet candidates are rejected if the reconstructed p_T is less than 6 GeV before jet energy scale corrections are applied (see below). This selection criterion is referred to below as the *jet reconstruction threshold*. Fake jet candidates arise mainly from noise in the calorimeter system, from electrons and from photons.

5.5.1 Jet Identification

The jet ID criterion used to reject candidates arising from electrons and photons is

- $EMfrac < 0.95$.

ID criteria used to remove jet candidates arising from noise are

- $0.05 < EMfrac$ (require confirmation from the EM calorimeter which is less sensitive to noise than the hadronic calorimeter);
- the fraction of the jet energy registered in the coarse hadronic layer should be below 40% (this layer is particularly sensitive to noise);
- the amount of energy detected by the L1 jet trigger must exceed either 80 GeV or 50% of the reconstructed jet energy (the trigger is an independent cross-check since its electronics is separate from the main read-out system).

The ID criteria given here are used for most detector regions. In some particular detector regions containing atypical instrumentation, like in the gap region between the CC and EC calorimeter, slightly modified criteria are used (see Ref. [95] for details).

5.5.2 Jet Identification Efficiency

The probability that a true jet gives rise to a jet candidate defines the jet *reconstruction efficiency*, $\epsilon_{\text{reco}}^{\text{jet}}$. If a jet candidate is found, the probability that it passes the jet ID criteria defines the jet *ID efficiency*, $\epsilon_{\text{ID}}^{\text{jet}}$. At low p_T , defined as below approximately 40 GeV, the measurement of $\epsilon_{\text{reco}}^{\text{jet}}$ is complicated due to the impact of the jet reconstruction threshold and the relatively large p_T resolution for jets: for example, a 20 GeV jet with a low π^0 content in the calorimeter shower (see Sect. 4.2.2) might fail the threshold cut. This cross-talk between reconstruction efficiency, reconstruction threshold and p_T resolution can be accounted for using the so-called *SSR* method (see Sect. 5.5.4). At large values of p_T , where the reconstruction threshold does not play a role, $\epsilon_{\text{reco}}^{\text{jet}}$ times $\epsilon_{\text{ID}}^{\text{jet}}$ reaches a *plateau* value which is approximately independent of p_T , and this plateau efficiency is measured using a tag-and-probe method [96, 97]. The presence and direction of flight of a jet is established by selecting a good jet or γ as tag. The probe jet is made up of one or more tracks with relatively high p_T which satisfy $\Delta\phi(\text{tag}, \text{tracks}) \approx \pi$. This selection is motivated by assuming a leading-order picture where the tag and the probe are back-to-back and have equal p_T . The validity of the leading-order picture is improved by rejecting events containing additional jets or a large amount of missing transverse energy. The jet reconstruction efficiency is measured as the probability that a jet candidate is found in the direction of the probe. Using a dijet sample, requiring a good jet as probe, gives a reconstruction efficiency in data which, as shown in Fig. 5.4, is $\sim 100\%$ above 40 GeV, while dropping down to $\sim 30\%$ at 15 GeV.

The jet ID efficiency corresponds to the probability that a reconstructed jet candidate passes the ID criteria given in Sect. 5.5.1, and it is found to be $\sim 95\%$ at 20 GeV, increasing to $\sim 99\%$ above 70 GeV (see Fig. 5.4). Simulation and data are in good agreement for $\epsilon_{\text{ID}}^{\text{jet}}$ over the whole p_T range, and for $\epsilon_{\text{reco}}^{\text{jet}}$ above 40 GeV. Residual disagreements on the 1 – 2% level are taken into account in the analysis by rejecting jets with a probability $\epsilon_{\text{reco}/\text{ID}}^{\text{data}}/\epsilon_{\text{reco}/\text{ID}}^{\text{sim}}$. The disagreements seen in $\epsilon_{\text{reco}}^{\text{jet}} < 40$ GeV are corrected for as part of the SSR method (see below). As a cross-check, the tag-and-probe method is also applied in a γ +jet sample, yielding efficiencies which are compatible with the dijet results within uncertainties. The γ +jet analysis has the advantage of a more precise measurement of the p_T of the probe, but it suffers from lower statistics and higher backgrounds.

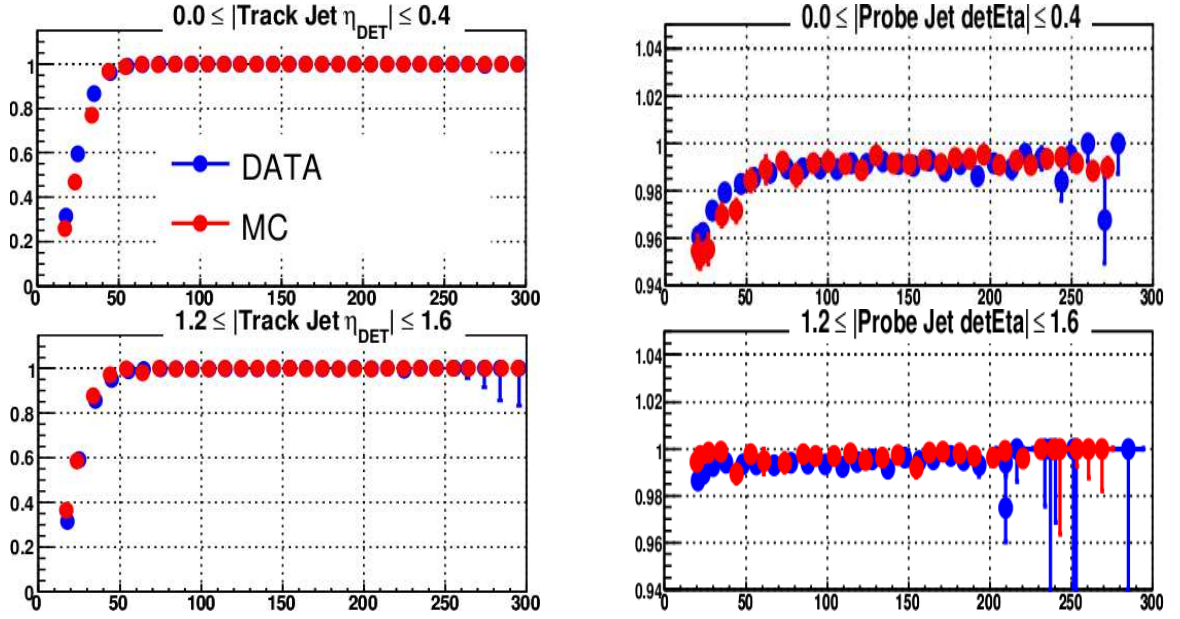


Figure 5.4: Jet reconstruction (left) and ID (right) efficiency measured in data and simulation using dijet events where the track(s) of the probe satisfies $0.0 < |\eta_{\text{det}}| < 0.4$ (upper row) or $1.2 < |\eta_{\text{det}}| < 1.6$ (lower row) (from Ref. [96]). The efficiencies are plotted as a function of the p_T of the tagged jet, as estimated from the p_T of the probe jet. No strong dependency on η_{det} was found.

5.5.3 Jet Energy Scale from γ +jet production

Jet energy scale (JES) corrections are applied to calibrate the reconstructed jet energies to, on average, be equal to the energy of the jet at the particle level. In $D\bar{O}$, the particle level is defined as the set of stable final-state particles of the interaction. Particles arising from interactions between the $p\bar{p}$ remnants (underlying event) from the hard scattering are included in the definition, whereas particles from additional $p\bar{p}$ interactions are not. The calorimeter energy associated with a reconstructed jet deviates from the particle-level jet energy due a number of effects which can be divided into three main categories [98]:

- The jet energy response, R , of the calorimeter is less than 1, mainly due to the lower energy response of the calorimeter to hadrons than to electromagnetic particles. Additional reasons are the presence of uninstrumented regions between modules as well as the effect of the zero suppression algorithm. The response is measured using the energy balance in γ +jet events. More specifically, the missing transverse energy component parallel to the axis defined by the direction of flight of the photon is assumed to arise from the calorimeter response. A veto is applied on sub-leading jets, and the γ and the jet are required to be back-to-back in the transverse plane. The response is found to be in the range $0.5 - 0.8$, depending on p_T^{jet} (see Fig. 5.5). This correction is the largest part of the JES corrections.
- The reconstructed jet energy receives a non-negligible contribution from other sources than the $p\bar{p}$ collision which produces the hard scattering. The resulting energy offset,

E_{offset} , arises from additional $p\bar{p}$ interactions from the same bunch crossing as the hard interaction, from noise in the calorimeter electronics, from decays of uranium in the calorimeter absorber plates and from pile-up of energy from previous bunch crossings. The size of E_{offset} is determined as a function of the number of reconstructed $p\bar{p}$ interaction vertices using data events from random bunch-crossings (zero bias events). The offset energies in events with 1-5 reconstructed primary vertices are shown in Fig. 5.5.

- Particles which contribute to the jet at particle level might not contribute to the detector-level jet, and the other way around. This can happen for instance if the particle fails to reach the calorimeter due to interactions with dead material or the magnetic field, or if energy from particles at the edge of the particle-level jet is deposited in calorimeter cells not belonging to the reconstructed jet. The net result of these effects referred to as detector showering, and is accounted for by a correction factor, F_{shower} , shown in Fig. 5.5. The shower correction is extracted from fitting the radial jet energy density profile in data, using templates derived in a simulated event sample by matching generated particles with energy deposits in the calorimeter, performed separately for particles included, and not included, in the particle-level jet.

In terms of the above-mentioned correction factors, the reconstructed, JES-corrected jet energy, E_{corr} , is defined as

$$E_{\text{corr}} \equiv \frac{E_{\text{cal}} - E_{\text{offset}}}{R \times F_{\text{shower}}} \quad (5.4)$$

To account for an insufficient accuracy in the simulation of the interaction between jets and the detector, the JES corrections are derived independently in data and simulation. Examples of the total JES correction factor for data, with uncertainties, are given in Fig. 5.6 as a function energy and η_{det} .

5.5.4 Jet Energy Scale and Resolution from Z +jet production

The use of Z +jet events to calibrate jet energies has a three-fold motivation. First, in contrast to the γ +jet and dijet samples, Z +jet offers a clean sample and a well-measured p_T of the tag object also at low p_T . Second, jets initiated by quarks and gluons have different response in the calorimeter due to the higher particle multiplicity in gluon-initiated jets. The fractions of quark and gluon-initiated jets are different in Z +jet events than in γ +jet events, and consequently jets in the two samples have a different average response. Third, as mentioned in Sect. 5.5.2, low- p_T jet samples are influenced by the jet reconstruction threshold and the jet energy resolution: low- p_T jets are only identified if their energy fluctuated upwards. To account for this, a method which simultaneously measures the energy scale and resolution has been devised, referred to as the *Shift, Smear and Remove* (SSR) method [99]. This method is defined after the γ +jet-derived JES corrections have been applied. The SSR method relies on the kinematic balance in Z +jet events where the Z and the jet are back-to-back in the transverse plane. By vetoing against softer jets it is established that $p_T^Z \sim p_T^{\text{jet}}$ on the particle level. The observable

$$\Delta S = \frac{p_T^{\text{jet}} - p_T^Z}{p_T^Z} \quad (5.5)$$

is studied in bins of p_T^Z , see Figs. 5.7 and 5.8. The ΔS spectrum is described by a Gaussian convoluted with an error function. The fit of the convoluted function to the ΔS distribution extracts the following information:

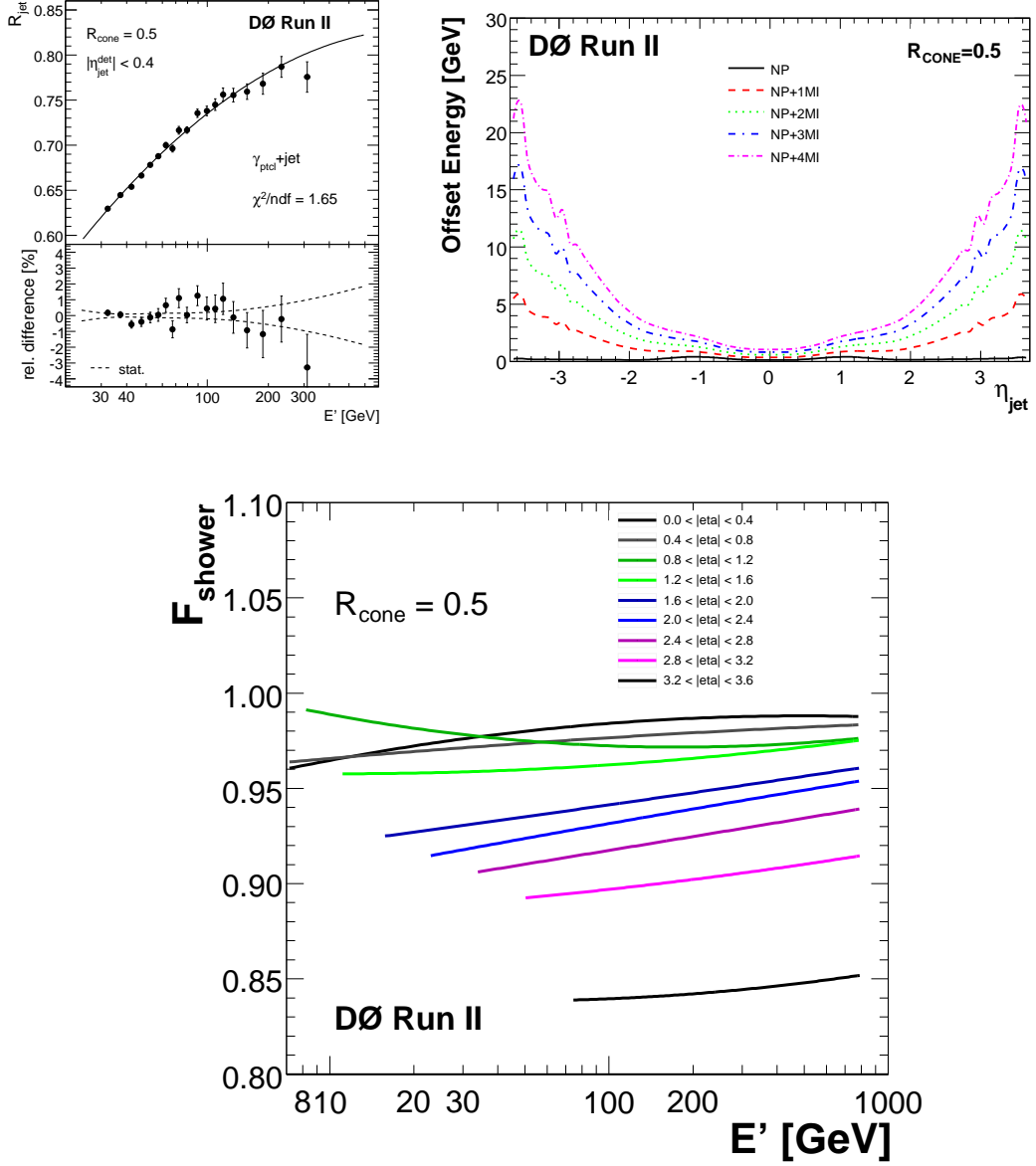


Figure 5.5: The three main steps of the DØ JES correction, as measured in data: the energy response for $|\eta_{\text{det}}^{\text{jet}}| < 0.4$ as a function of the true jet energy E' estimated from the photon energy (top, left); the energy offset as a function of η_{det} for events with and the number of reconstructed interactions vertices (top, right); and the detector showering correction as function of the estimated true jet energy and pseudorapidity (bottom) [98]. For the energy offset, separate curves are shown as function of the number of reconstructed interaction vertices. For zero vertices the reconstructed energy is assumed to arise from noise and pile-up from previous bunch crossings (labelled NP). For events with n interaction vertices the curve is labelled NP+ n MI, with MI meaning multiple $p\bar{p}$ interactions.

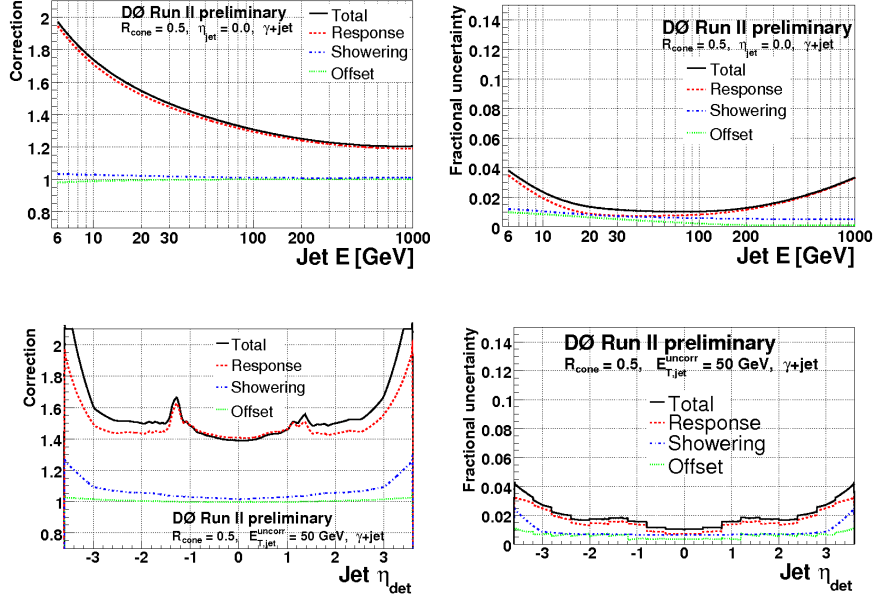


Figure 5.6: The total jet energy scale correction for data (left) and the corresponding uncertainties (right) shown as a function of the uncorrected jet energy for $\eta_{\text{det}} = 0$ (top) and as a function of η_{det} for $E_{\text{jet}}^{\text{uncorrected}} = 50$ GeV (bottom) [98].

- the p_T^Z value corresponding to the peak of the Gaussian, $\langle \Delta S \rangle$, measures the JES;
- the width of the Gaussian, $\sigma_{\Delta S}$, measures the jet energy resolution (JER);
- the turn-on p_T value where of the error function measures the p_T^{jet} value which corresponds to the $E_{\text{jet}}^{\text{uncorrected}} = 6$ GeV reconstruction threshold.

The fits are performed independently for data and simulation. The $\langle \Delta S \rangle$ and $\sigma_{\Delta S}$ distributions are given as a function of p_T^Z in Fig. 5.7. For data-events with $p_T^Z = 20$ GeV, the jet has an average particle-level p_T of ~ 15.4 GeV ($\langle \Delta S \rangle \approx -0.3$), against 16.7 GeV in simulation ($\langle \Delta S \rangle \approx -0.2$). This deviation probes the difference in jet energy response between data and simulation, and the reconstructed jet energies in simulation are shifted so that agreement with data is seen for the $\langle \Delta S \rangle$ distribution. The $\sigma_{\Delta S}$ in data is larger than in simulation, and it is assumed that this is due to differences between the GEANT-based detector simulation and the real DØ detector. For instance, the superconducting solenoidal coil structure is described in the simulation as a homogeneous, ϕ -symmetric cylinder [72]. In reality, some particles pass through the solenoidal system without intersecting the coil, i.e. passing through less material than in the simulation before reaching the calorimeter. In contrast, those particle which pass through the coil are exposed to significantly more material than in the simulation. This contributes to explain why the energy resolution is larger in data than in simulation. To bring the $\sigma_{\Delta S}$ distribution in simulation into agreement with data, reconstructed jet energies in simulation are smeared with a Gaussian function with a width given as $\sigma_{\text{smeared}}^2 = \sigma_{\Delta S, \text{data}}^2 - \sigma_{\Delta S, \text{sim}}^2$.

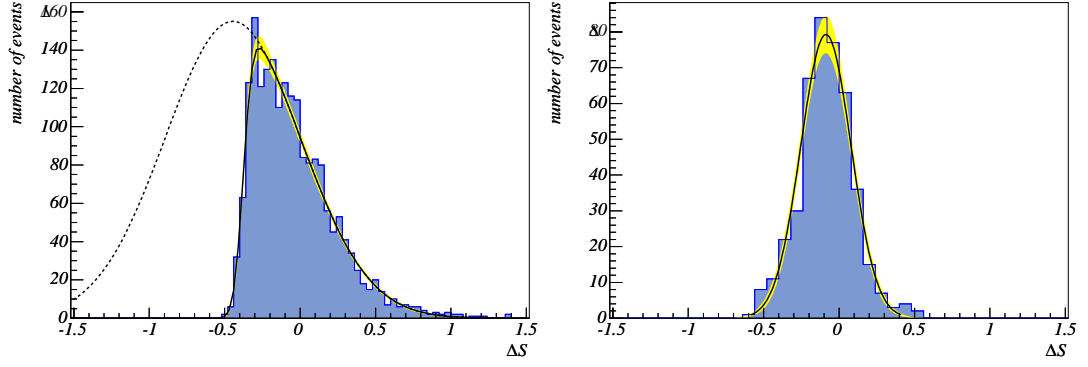


Figure 5.7: The ΔS distribution for $18 < p_T^\gamma < 23$ GeV (left) and for $70 < p_T^\gamma < 75$ GeV (right) as reconstructed from γ +jet events in data, fitted with the convolution of a Gaussian and an error function (continuous line) [100]. The Gaussian part of the fit is given as a dotted line. These plots are shown for γ +jet production since the corresponding plots for Z/γ^* production were not available. The qualitative properties of the ΔS distribution are identical for Z/γ^* +jet and γ +jet production.

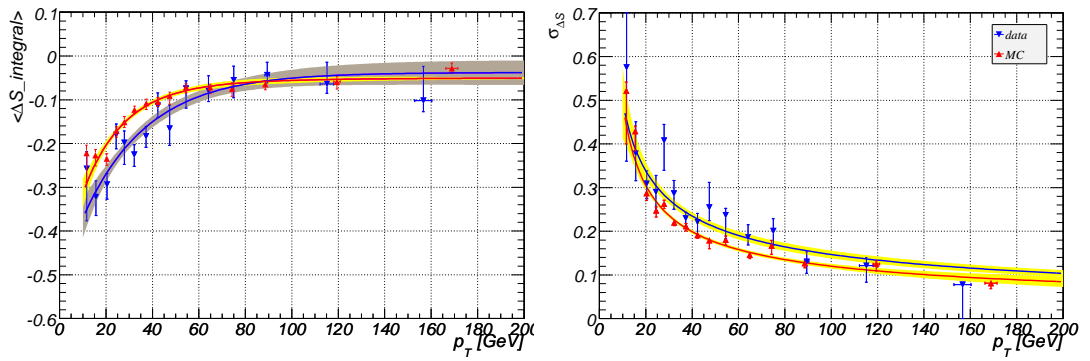


Figure 5.8: The distributions of $\langle \Delta S \rangle$ (left) and $\sigma_{\Delta S}$ (right) extracted from Z/γ^* +jet events in data (blue) and simulation (red), given as functions of p_T^Z [99].

Chapter 6

Event Samples from Data and Simulation

In this Chapter, a description of the data and simulated event samples which are used in the analysis presented in this thesis is given.

6.1 Data Sample

The analyzed data sample was recorded by the DØ experiment between October 2002 and February 2006 during Run II of the Tevatron accelerator. In order to reduce the number of events which needs to be taken into account in each DØ analysis, sub-samples (*skims*) of the full data set which satisfy very loose selection criteria have been compiled by the Common Sample Group [101]. In this analysis a skim containing all events where at least two EM calorimeter objects satisfying $p_T > 12$ GeV was used.

6.1.1 Data Quality

The analysis presented in this thesis relies on the accurate identification of electrons and jets and a precise measurement of their kinematic properties. To guarantee that the detector was operating properly during the time period when the analyzed data sample was registered, so-called *data quality selection criteria* are employed. In addition to rejecting events where the tracking, calorimeter or luminosity measurement systems were malfunctioning, jet candidates which contain specific calorimeter cells known to have a high level of noise are discarded.

6.1.2 Integrated Luminosity

The integrated luminosity has been determined from the rate of inelastic $p\bar{p}$ interactions measured using the luminosity monitoring system (see Sect. 4.2.4), yielding

$$\mathcal{L} = \int L dt = 1.04 \text{ fb}^{-1}, \quad (6.1)$$

with a relative uncertainty of 6.1% [102].

6.2 Simulated Event Samples

Simulated event samples are first used for to comparing the predictions of the PYTHIA v6.314 and SHERPA 1.0.6 event generator models with kinematic distributions reconstructed in the data sample in Chapter 7. Thereafter, event samples simulated using PYTHIA v6.314 and ALPGEN v2.05 + PYTHIA v6.325 are used to estimate background contributions and the impact of the detector on the reconstructed kinematic distributions in data, thereby correcting the

distributions to the particle level. The configuration parameters of PYTHIA were set according to Tune A [51]. Each event was passed through a GEANT-based [77] simulation of the DØ detector [78]. Whereas the PYTHIA and ALPGEN+PYTHIA samples were provided by the Common Samples Group, the SHERPA sample was generated specifically for these studies. An overview of the simulated event samples used in the studies and measurements presented in chapters 7 and 9 is given in Table 6.1. The ALPGEN+PYTHIA samples were generated in six different bins depending on the number of partons included in the matrix-element calculation. The quoted cross section for each multiplicity bin is used when merging the samples to build a jet-inclusive sample. For SHERPA, up to three partons were included in the matrix-element calculations. All $Z/\gamma^* \rightarrow e^+e^-$ samples were generated with the following requirement¹: $60 < M_{Z/\gamma^*} < 130$ GeV.

Process	Generator	$\sigma \times \text{BR}$	Events
$Z/\gamma^* \rightarrow e^+e^- + 0$ partons	ALPGEN+PYTHIA	139.0 pb	2.23M
$Z/\gamma^* \rightarrow e^+e^- + 1$ parton	ALPGEN+PYTHIA	42.0 pb	808k
$Z/\gamma^* \rightarrow e^+e^- + 2$ partons	ALPGEN+PYTHIA	10.2 pb	400k
$Z/\gamma^* \rightarrow e^+e^- + 3$ partons	ALPGEN+PYTHIA	2.4 pb	202k
$Z/\gamma^* \rightarrow e^+e^- + 4$ partons	ALPGEN+PYTHIA	0.49 pb	100k
$Z/\gamma^* \rightarrow e^+e^- + 5$ partons	ALPGEN+PYTHIA	0.12 pb	50k
$Z/\gamma^* \rightarrow e^+e^- + \leq 3$ partons	SHERPA		3M
$Z/\gamma^* \rightarrow e^+e^-$	PYTHIA		1.2M
$Z/\gamma^* \rightarrow \tau^+\tau^- + 0$ partons	ALPGEN+PYTHIA	139.0 pb	1.4M
$Z/\gamma^* \rightarrow \tau^+\tau^- + 1$ parton	ALPGEN+PYTHIA	42.0 pb	400k
$Z/\gamma^* \rightarrow \tau^+\tau^- + 2$ partons	ALPGEN+PYTHIA	10.2 pb	200k
$Z/\gamma^* \rightarrow \tau^+\tau^- + 3$ partons	ALPGEN+PYTHIA	2.4 pb	200k
$Z/\gamma^* \rightarrow \tau^+\tau^-$	PYTHIA		1.2M
$W \rightarrow e\nu$	PYTHIA		4.2M
WW	PYTHIA		200k
$t\bar{t}$	PYTHIA		1.1M

Table 6.1: Information on the simulated event samples used in chapters 7 and 9.

¹The contribution to the selected event samples in the presented analysis stemming from other ranges in M_{Z/γ^*} was found to be negligible.

Chapter 7

Detector-Level Studies of $Z/\gamma^* + \text{jets}$ production

An accurate understanding of the associated production of Z/γ^* and jets is of great importance to hadron-collider experiments. In addition to representing an important test of perturbative QCD, this channel is a major source of background events to searches for new phenomena, like SUSY and Higgs production, as well as to studies of known, rare Standard Model processes, e.g. di-boson or single-top production. As described in Chapter 3, a wide range of algorithms for describing QCD radiation has become available over the last years. Several different algorithms for consistently combining matrix-element and parton-shower descriptions of QCD radiation have been presented and implemented in event generator codes, e.g. SHERPA and ALPGEN+PYTHIA, yielding predictions on the particle-level suitable for a detailed detector simulation. Although these recent advances are not applicable only for $Z/\gamma^* + \text{jets}$ production, this channel offers the advantage of a large cross section and a clean experimental signature. Experimental measurements of jet-related properties in the $Z/\gamma^* + \text{jets}$ channel therefore offers an ideal framework for validating, comparing and improving different event generator codes. In this Chapter the properties of jets in $p\bar{p} \rightarrow Z/\gamma^*(\rightarrow e^+e^-) + \text{jets}$ production are reconstructed in data and compared with the predictions of SHERPA and PYTHIA.

7.1 Event Selection

The event selection starts by requiring two electron candidates of opposite charge, both having a p_T above 25 GeV and being reconstructed within the pseudorapidity range $|\eta_{\text{det}}| < 1.1$ or $1.5 < |\eta_{\text{det}}| < 2.5$. At least one of the electrons is required to be reconstructed in the central part of the detector ($|\eta| < 1.1$). Both electron candidates must satisfy a likelihood criterion optimized for separating real electron from backgrounds mainly arising from jet production via QCD processes (see Sect. 5.4.1). The di-electron invariant mass (M_{ee}) must satisfy $70 < M_{ee} < 110$ GeV, and the event must have been triggered by a single-electron or di-electron trigger. It is assumed that the trigger efficiency is 100%, and also that all events in the selected data sample correspond to true $Z/\gamma^* \rightarrow e^+e^-$ events, i.e. that there is zero background¹. Jets were defined using the DØ Run II Mid-Point algorithm (see Sect. 3.7.1) and the jet identification criteria given in Sect. 5.5. Only jets satisfying $p_T > 15$ GeV and $|\eta_{\text{det}}| < 2.5$. In addition, each jet must have a separation in ΔR to all identified electron candidates larger than 0.5.

A total of 50,440 data events pass the event selection. Of these, 40.437/7.976/1.629/332/52 events have 0/1/2/3/4 reconstructed jets with p_T^{jet} above 15 GeV. The properties of the se-

¹These assumptions will be justified in more detailed studies presented Chapt. 9.

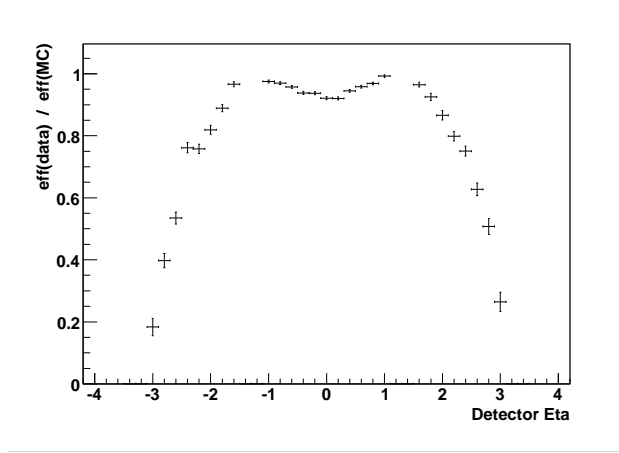


Figure 7.1: The identification correction weight attached to each electron in the simulated sample, as a function of η_{det} [88].

lected a $Z/\gamma^* \rightarrow e^+e^-$ event sample is compared to two event samples which were generated using the PYTHIA and SHERPA event generators and passed through a GEANT-based simulation of the DØ detector (see Sect. 6.2).

7.2 Correcting the Performance of the Simulated Detector

The properties of reconstructed distributions in simulation are determined by the physics model of the event generator and of the performance of the simulated detector. The purpose of this study is to use DØ data to test and validate the physics models of SHERPA, using PYTHIA as a reference. To achieve this, the performance of the simulated detector must be identical to that of the real detector within the quoted uncertainties. This is done by measuring the electron and jet identification efficiencies, energy scales and energy resolutions in data and simulation, and then by modifying the simulated event sample to account for any discrepancies.

As noted in Sect. 5.4.2, the electron identification efficiency in simulation is higher than in data. To compensate for this, each simulated event which passes the event selection is given a weight

$$\epsilon^{\text{corr}}(\eta_{\text{det},1}) \cdot \epsilon^{\text{corr}}(\eta_{\text{det},2}), \quad (7.1)$$

where $\eta_{\text{det},1}$ and $\eta_{\text{det},2}$ are the η_{det} -values of the two electrons and ϵ^{corr} is the ratio of the electron identification efficiency measured in the sample and the simulation sample using the tag-and-probe method, as a function of η_{det} , as given in Fig. 7.1.

The reconstructed jet energies were calibrated using jet energy scale corrections derived in a γ -jet sample [103]. To measure the differences in jet energy scale and resolution between data and simulation the sub-sample of the $Z/\gamma^* \rightarrow e^+e^-$ sample containing one and only one jet and satisfying $\Delta\phi(\text{di-electron, jet}) > 2.9$ was selected (see Sect. 5.5.4). The mean and RMS of the distribution of

$$\Delta S = \frac{p_T^{\text{jet}} - p_T^Z}{p_T^{\text{jet}}} \quad (7.2)$$

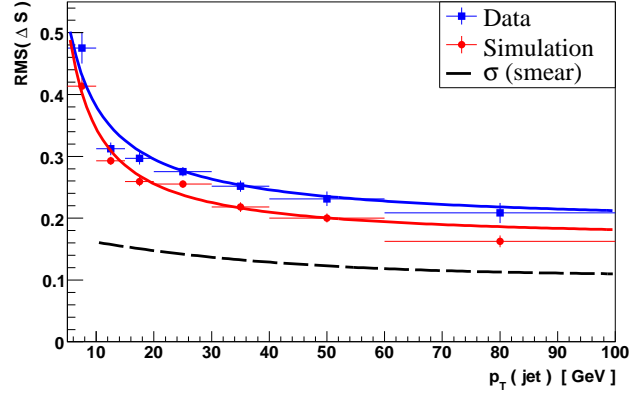


Figure 7.2: The measured jet energy resolution in data and simulation, fitted with Eq. (7.3). The difference in quadrature gives the width ($\sigma(\text{smear})$) of the Gaussian function with which reconstructed jet energies in simulation were smeared to remove the discrepancy.

were studied in bins of p_T^Z . The mean, which is sensitive to the jet energy scale, in data and simulation was found to be compatible within uncertainties, indicating that the applied jet energy scale version is valid within uncertainties also for $Z/\gamma^* + \text{jet}$ events. The RMS distribution in simulation was found to be smaller than in data, see Fig. 7.2. The two RMS distributions were fitted with the standard expression for relative energy resolution,

$$\left(\frac{\Delta p_T}{p_T}\right) = \sqrt{A^2 + \frac{B^2}{p_T} + \frac{C^2}{p_T^2}}. \quad (7.3)$$

The first term, A , is independent of p_T and reflects inhomogenities in the calorimeter, whereas the second term (B) arises from the the statistical uncertainty in the calorimeter energy sampling. Noise in the calorimeter and the read-out system is described by the third term (C). As can be seen in Fig. 7.2, the jet energy resolution is smaller in simulation than in data. To correct for this, reconstructed jet energies in the simulated event sample were smeared using a Gaussian function with a width given by $\sigma_{\text{corr}} = \sqrt{(f_{\text{data}}^{\text{JER}}(p_T))^2 - (f_{\text{sim}}^{\text{JER}}(p_T))^2}$.

In Fig. 7.3 the invariant-mass distributions (M_{ee}) in PYTHIA and SHERPA are compared with data. The PYTHIA spectrum is shifted towards higher values of M_{ee} with respect to data. The same tendency is seen for SHERPA, but the shift is slightly larger than for PYTHIA. These discrepancies are in contrast to the fact that the electron energy scale is deduced by requiring the M_{ee} distribution in simulation to match that in data (see Sect. 5.4.3). In the case of PYTHIA, the discrepancies can be traced back to the simulation of the energy of electrons which deposite their energy close to the boundary between two calorimeter towers², referred to as non-fiducial electrons. By selecting only events where both electrons deposited their energy well separated from from a ϕ -crack, referred to as fiducial electrons, agreement between PYTHIA and data is seen, as illustrated in Fig. 7.4. Since the azimuthal angles of the electrons do not correlate with any of the observables of interest, the mismatch between

²Regions referred to as ϕ -cracks, see Sect. 4.2.2

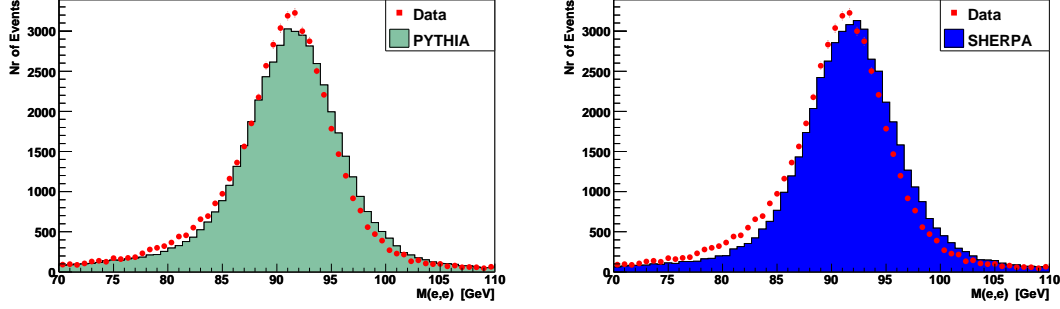


Figure 7.3: The distribution of the di-electron invariant mass ($M(e,e)$) in data compared with in PYTHIA (left) and with SHERPA (right).

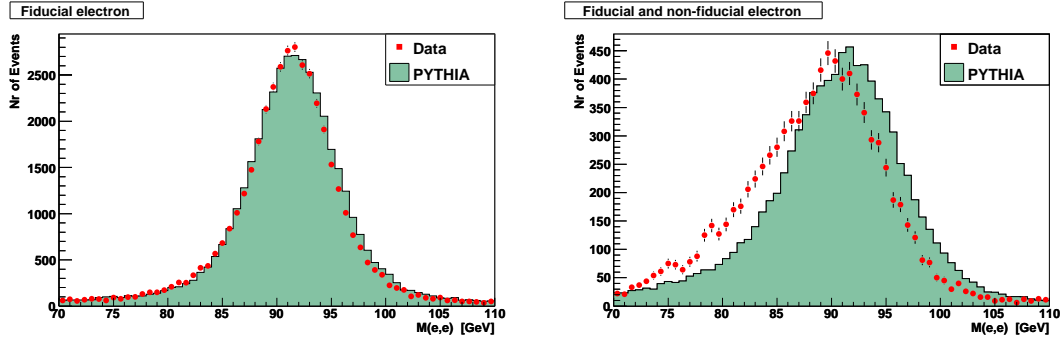


Figure 7.4: The distribution of the di-electron invariant mass in data compared with in PYTHIA for events where both electrons passed through a well-instrumented (fiducial) region of the detector (left). The same distribution, but for events where one of the electrons passed through a boundary (non-fiducial) region between two calorimeter towers (right).

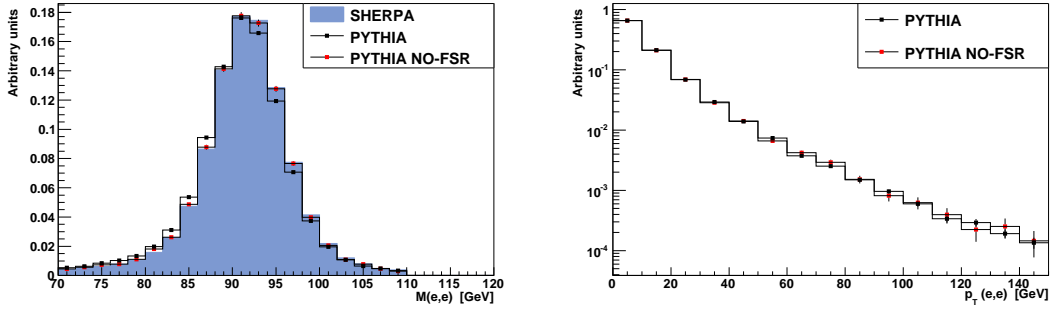


Figure 7.5: The di-electron invariant mass distribution ($M(e, e)$) for SHERPA, the inclusive PYTHIA sample and for the PYTHIA sub-sample of events where no $e \rightarrow e\gamma$ splitting occurred, labelled *NO-FSR* (left). The di-electron transverse momentum distribution ($p_T(e, e)$) for the two PYTHIA samples (right).

data and simulation arising from the inclusion of non-fiducial electrons in the analysis is not expected to influence the analysis significantly. The deviations between the M_{ee} spectra in SHERPA and PYTHIA reflect differences in the simulation of higher-order QED corrections to the boson decay. In PYTHIA, the boson is first allowed to decay into an e^+e^- pair, then a QED parton-shower algorithm is applied to simulate $e \rightarrow e\gamma$ splittings (*QED FSR*). The SHERPA version used for this part of the analysis did not model QED corrections to the boson decay, and as a result the electrons in the SHERPA sample have a higher average p_T and M_{ee} than in PYTHIA, as seen in Fig. 7.5. To verify that the difference between the two simulations is entirely due to QED FSR the PYTHIA sub-sample consisting of events where no $e \rightarrow e\gamma$ splitting took place was selected, and as seen in Fig. 7.5 (left), the M_{ee} distribution of this PYTHIA sub-sample reproduces the SHERPA distribution³. The decay mode of the Z/γ^* is expected to be uncorrelated with the jet activity in the event. To test whether the p_T^Z distribution in the SHERPA sample is influenced by the lack of a QED FSR model the p_T^Z distributions in the inclusive PYTHIA sample and the sub-sample without $e \rightarrow e\gamma$ splittings were compared, and agreement was seen within statistical uncertainties (Fig. 7.5(right)).

7.3 Detector-level Comparisons: Data, PYTHIA and SHERPA

With agreement between the simulated detector and the real detector any remaining differences between data and simulation must be due to the physics model of the event generator. In Fig. 7.6 the p_T distribution of the di-electron system measured in data is shown and compared to the predictions of PYTHIA (left) and SHERPA (right). The indicated ranges for each generator reflect the predicted value and the statistical uncertainty of the simulated event sample. In the lower part of each plot the ratio of the number of events in data to that predicted by the simulation is given. An upward slope, corresponding to a lack of events containing a di-electron system with a large p_T value, is found in the PYTHIA event sample. This indicates a lack of events containing a high- p_T jet. For SHERPA, the agreement in the

³Initiated by these studies a model for simulating higher-order QED corrections was added to later versions of SHERPA.

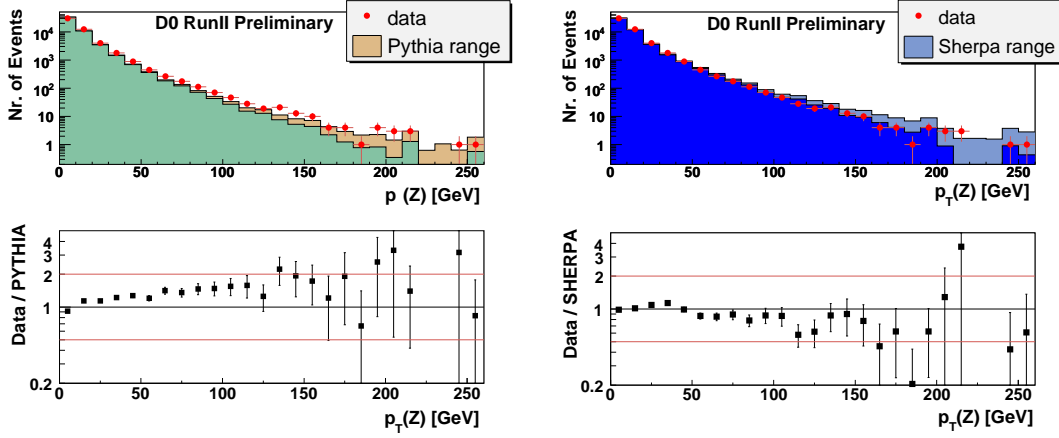


Figure 7.6: The distribution of the transverse momentum of the Z boson ($p_T(Z)$) in data and PYTHIA (left), and in data and SHERPA (right). The lower plots show the ratio of the distribution in data to in the simulation. The red lines indicate a factor two up and down.

low p_T range is reasonable, but the simulated sample contains more events with a large value of p_T than seen in data.

The jet multiplicities observed in data and predicted by the two event generator models are given in Table 7.1. The simulated event samples were normalized to the total number of events observed in data, with no separate normalization is performed for the various jet multiplicity bins. The SHERPA prediction is in good agreement with the jet multiplicities observed in data. In contrast, PYTHIA show deviations which increase with the jet multiplicity. However,

Table 7.1: Number of events for different jet multiplicities measured in data compared with the predictions of PYTHIA and SHERPA.

Sample	Inclusive	0-jet	1-jet	2-jet	3-jet	4-jet
Data	50417	40624	7877	1552	306	52
SHERPA	50417	39746	8410	1842	335	58
PYTHIA	50417	41271	7604	1324	193	23

as seen in Fig. 7.7, the predictions of both event generators are in agreement with data within the large systematic uncertainties arising from low p_T jets, which dominate the distributions. The uncertainty for the ratio between data and each simulated prediction is evaluated treating the jet energy scale error of data and simulation as uncorrelated. Given that data and simulation share common sources of uncertainty, this is a conservative estimate.

The differential cross section $d\sigma/dp_T$ for the leading jet⁴ (Fig. 7.8), is consistent with the indications from the $p_T(Z)$ spectra, with a positive slope observed in the ratio for the PYTHIA

⁴The jets are ordered in terms of decreasing p_T^{jet} .

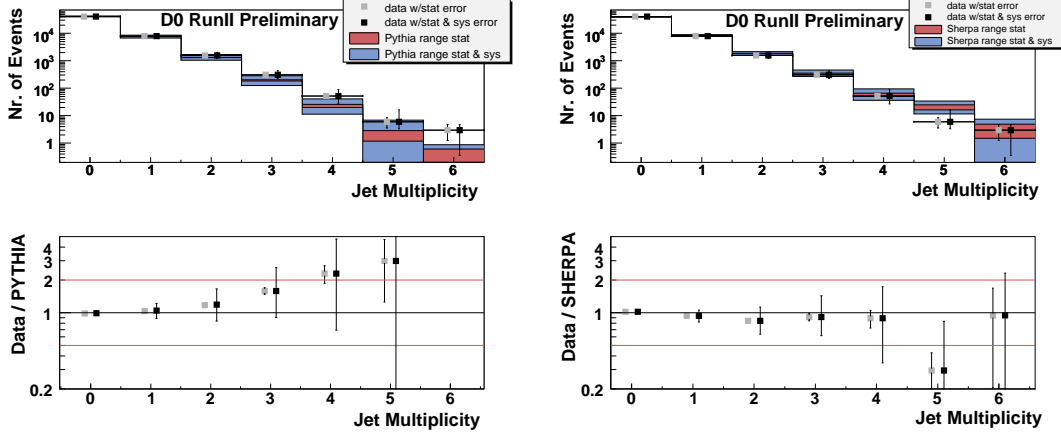


Figure 7.7: The jet-multiplicity distribution in data and PYTHIA (left), and in data and SHERPA (right). The lower plots show the ratio of the distribution in data to in the simulation. The red lines indicate a factor two up and down.

prediction. The slope is found to be larger for the second and third leading jets (Figs. 7.9 and 7.10). For example, a factor of 5 more events containing at third leading jet with $p_T \approx 50$ GeV is seen in data than is predicted by PYTHIA. The SHERPA prediction for the p_T of the leading jet, Fig. 7.8(right), is consistent with data within the systematic uncertainties in most bins. The largest deviations, apart from the highest p_T bin where statistics is small, are found at around 80 GeV, where SHERPA predicts a factor 1.3 more jets than seen in data. The p_T spectra for the second and third leading jets, Figs. 7.9 and 7.10, show comparable level of agreement between SHERPA and data as for the leading jet. The largest deviations for both the second and third leading jets are seen at around 80 GeV where SHERPA predicts a factor 1.7 more jets than seen in data.

Both event generators offer a good description of the difference in pseudorapidity between the two leading jets, $\Delta\eta(\text{jet}, \text{jet})$, evaluated in events with two or more jets, as seen in Fig. 7.11. The distributions for the corresponding difference in the azimuthal angle, $\Delta\phi(\text{jet}, \text{jet})$, are shown in Fig. 7.12. SHERPA gives a good description of data, and ratio of SHERPA to data agrees well with unity within the uncertainties. Also the PYTHIA prediction for the shape of $\Delta\phi(\text{jet}, \text{jet})$ distribution agrees well with data except at $\Delta\phi = \pi$ where a significant peak is seen for the PYTHIA prediction, which is not observed in data. As for SHERPA, the overall normalization agrees with data.

An interesting class of events are those with three jets satisfying

$$|\eta_1 - \eta_2| > 2.0 \quad (7.4)$$

and

$$\eta_1 < \eta_3 < \eta_2 \quad \text{or} \quad \eta_2 < \eta_3 < \eta_1 \quad (7.5)$$

where η_i is η of the i^{th} leading jet in the event. The motivation for studying these events is the similarity to the vector boson fusion production channel of the Higgs boson. At parton-level this channel is characterized by one outgoing quark in each forward direction and a Higgs particle decaying in the central part of the detector. Due to the Higgs being color neutral, a

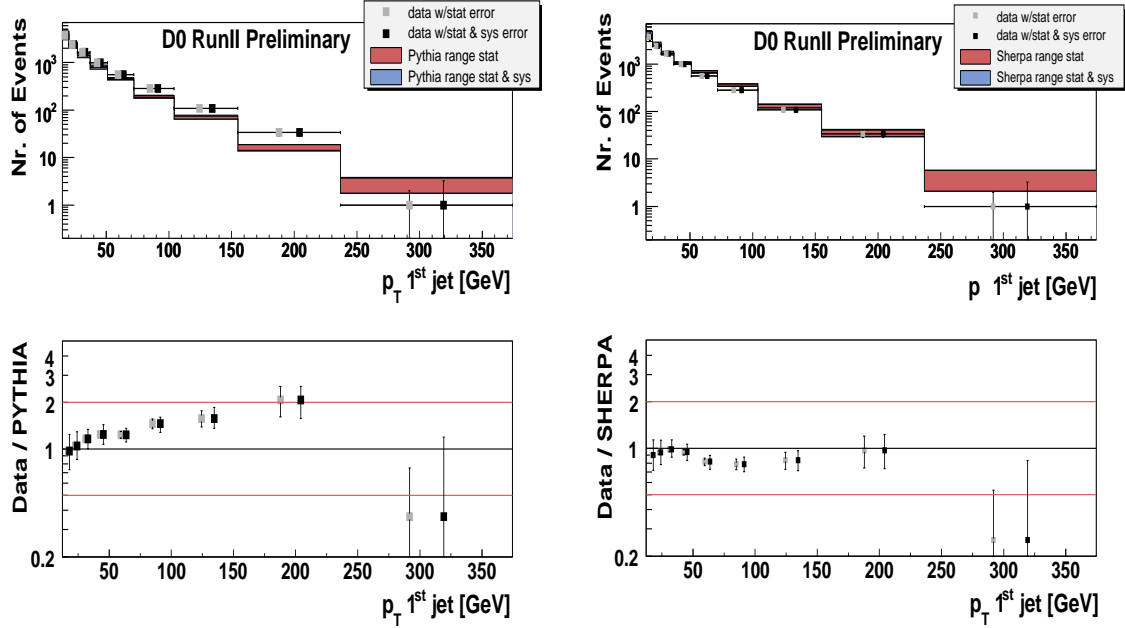


Figure 7.8: The p_T distribution of the leading jet in data and PYTHIA (left), and in data and SHERPA (right). The lower plots show the ratio of the distribution in data to in the simulation. The red lines indicate a factor two up and down.

low level of QCD activity in the central detector region is expected. A veto on the jet activity in the central part of the detector will be a powerful tool to suppress the large $t\bar{t}$ background, but such a veto can only be applied if its efficiency is known for all contributing processes, and this requires an accurate modelling of the associated production of QCD jet and bosons.

The p_T distributions of the third leading jet in events satisfying (7.4) and (7.5) are shown in Fig. 7.13. The description of η of the third jet, relative to the two leading jets, is given by

$$\eta^* = \eta_3 - \frac{\eta_1 + \eta_2}{2}, \quad (7.6)$$

which is shown in Fig. 7.14. The number of events passing the tagging criteria is rather low, but the SHERPA prediction is compatible both with the overall rate and the shape of the distribution observed in data. PYTHIA predicts a factor 1.7 fewer 3-jet events than seen in data, and therefore also gives too few events passing the tagging criteria. However, the shape of the distribution is correctly described.

7.4 Conclusions

A comparison between $D\bar{O}$ data and the event generator SHERPA for the associated production of $Z/\gamma^* \rightarrow e^+e^-$ and n jets has been presented, including a PYTHIA sample as reference. The PYTHIA simulation describes jets through a parton shower algorithm which has been tuned to match a matrix element prediction for $Z/\gamma^* + 1\text{-jet}$ production. The p_T spectra of the di-electron system and for the three leading jets show that PYTHIA predicts fewer jets with a large value of p_T than seen in data, and that the discrepancy increases with jet multiplicity.

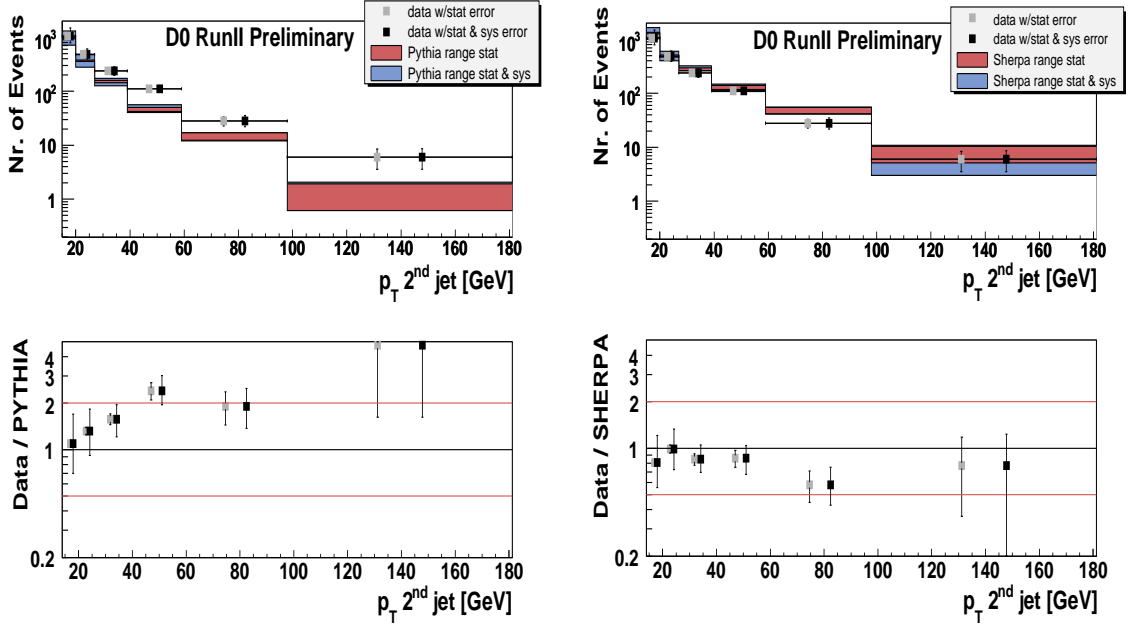


Figure 7.9: The p_T distribution of the second jet in data and PYTHIA (left), and in data and SHERPA (right). The lower plots show the ratio of the distribution in data to in the simulation. The red lines indicate a factor two up and down.

In the 2-jet sample, $\Delta\eta(\text{jet}, \text{jet})$ is well described. In the $\Delta\phi(\text{jet}, \text{jet})$ distribution a significant peak at π is seen in the PYTHIA sample but not in data. In contrast, SHERPA event generator, combining parton shower and matrix element description of jets using the CKKW algorithm, has been found to offer a good description of the studied jet properties. SHERPA has been shown to give an accurate description of jet multiplicities up to four jets. The p_T spectra predicted by SHERPA for the di-electron system as well as for the leading, second and third highest p_T jets are in reasonable agreement with the spectra observed in data. Also angular correlations between the two leading jets in events with two or more jets are well described by SHERPA.

7.5 Limitations of Detector-level Comparisons

To illustrate the need for measurements corrected for detector effects, comparisons between the predictions of SHERPA v1.0.6 and SHERPA v1.0.8 are shown in Fig. 7.15, with the total, jet-inclusive number of events in each sample being equal. The newer SHERPA version predicts 10 – 20% less leading jets with a p_T above ~ 60 GeV than the older version. The ratio of the prediction of v1.0.6 to data in Fig. 7.8 indicate indirectly that the newer version is in good agreement with data. However, the newer version predicts significantly less 3-jet events than v1.0.6, whereas data and v1.0.6 agree (see Fig. 7.9). Clearly, the possibility to compare new versions of SHERPA and other generators directly with data would be of great importance, and to facilitate such comparisons the reconstructed distributions in data must be corrected for the impact of the detector. In the following two Chapters, the p_T^{jet} spectra and angle $\Delta\phi(\text{di-}$

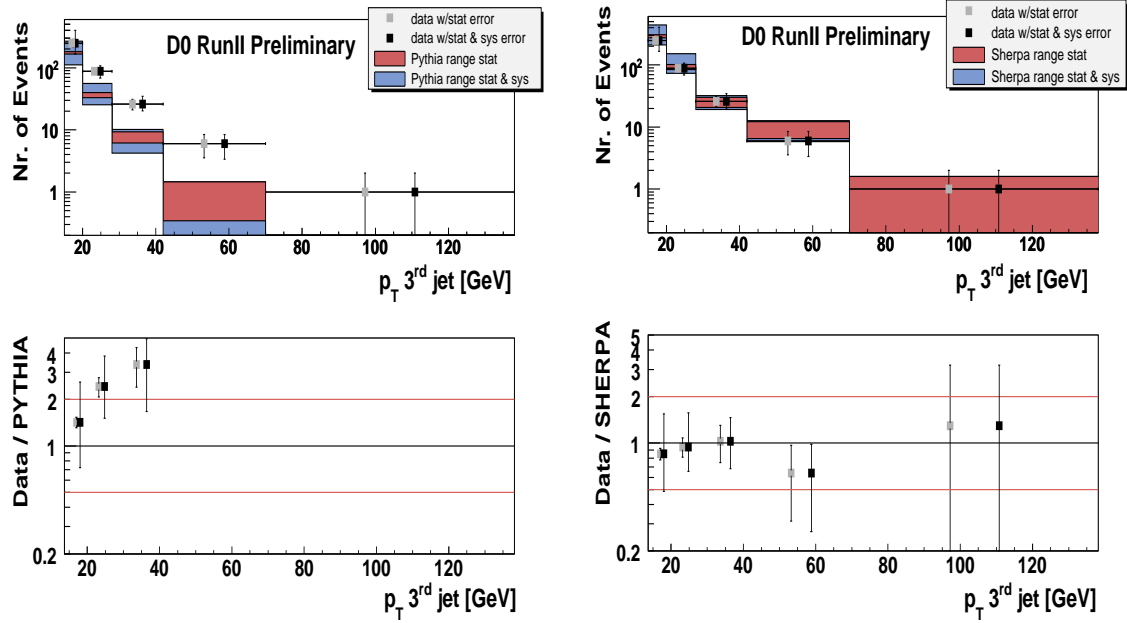


Figure 7.10: The p_T distribution of the third jet in data and PYTHIA (left), and in data and SHERPA (right). The lower plots show the ratio of the distribution in data to in the simulation. The red lines indicate a factor two up and down.

electron, jet) in $Z/\gamma^*(\rightarrow e^+e^-) + N$ jets data from DØ are measured and used as a benchmark for detailed comparisons between various commonly used event generator models.

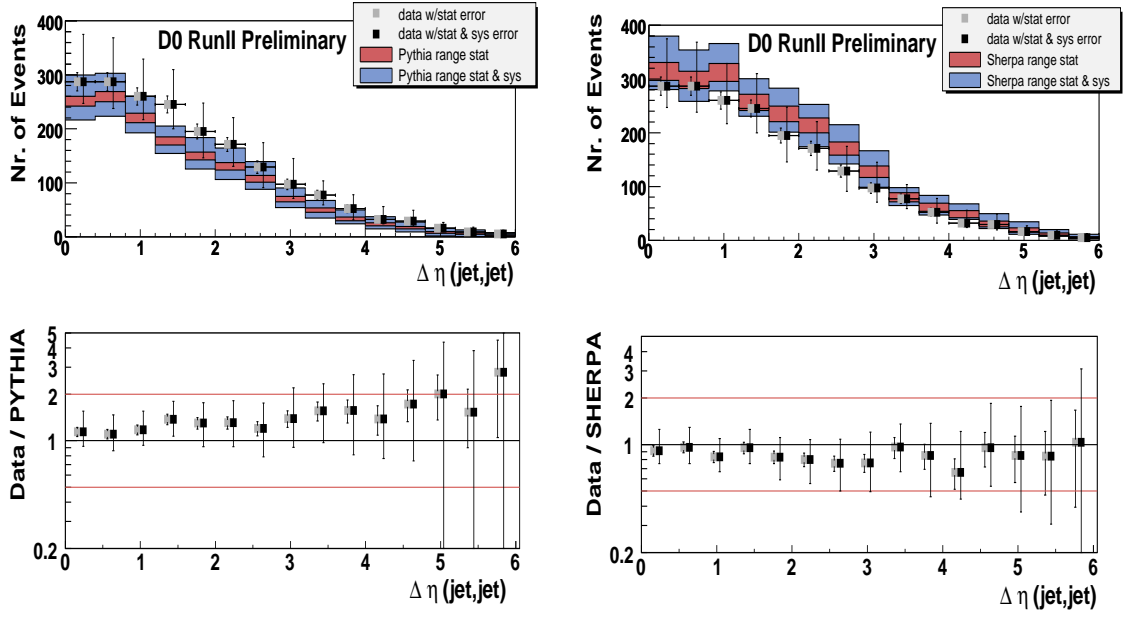


Figure 7.11: The distribution of the difference in pseudorapidity between the two leading jets in data and PYTHIA (left), and in data and SHERPA (right). The lower plots show the ratio of the distribution in data to in the simulation. The red lines indicate a factor two up and down.

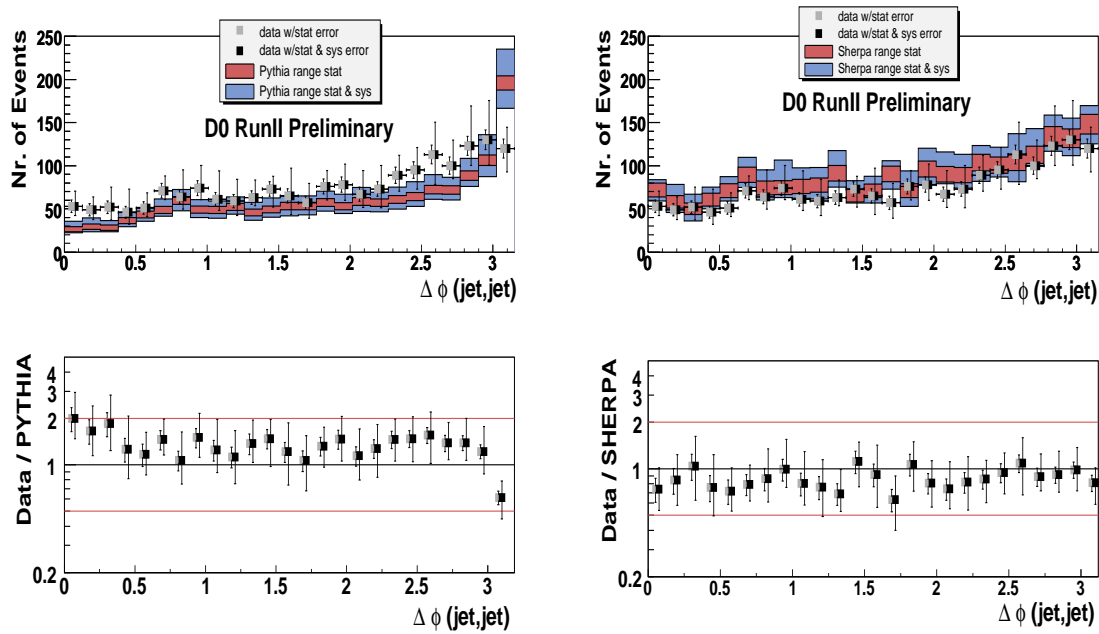


Figure 7.12: The distribution of the difference in azimuthal angle between the two leading jets in data and PYTHIA (left), and in data and SHERPA (right). The lower plots show the ratio of the distribution in data to in the simulation. The red lines indicate a factor two up and down.

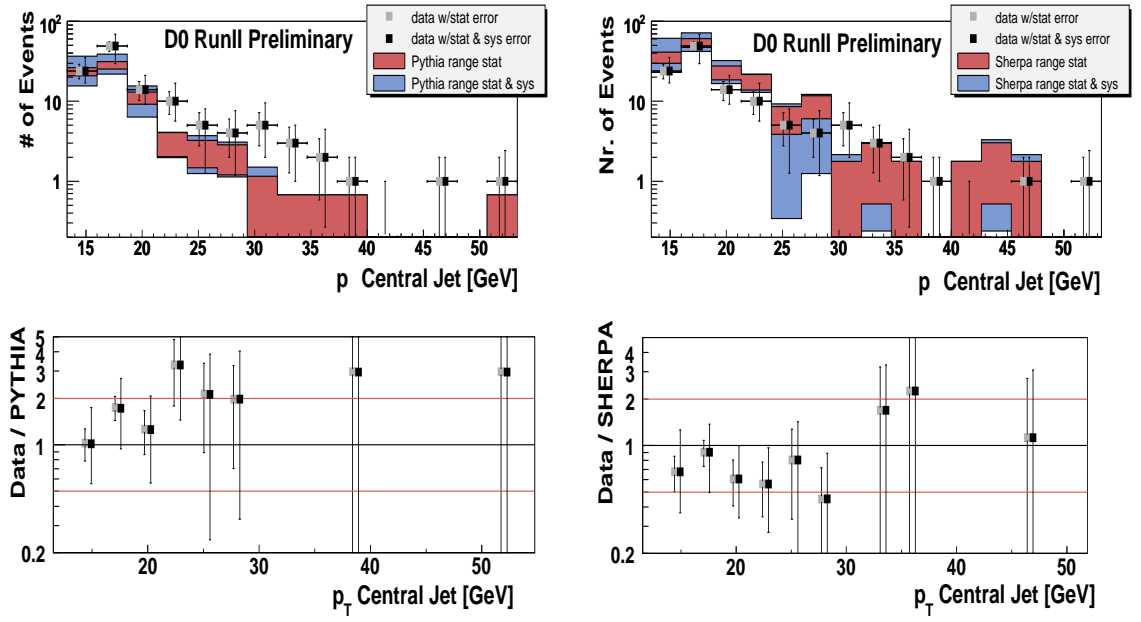


Figure 7.13: Events containing three jets were selected if the the difference in pseudorapidity between the two leading jets exceeds 2.0. The p_T of the third jet is shown for data and PYTHIA (left), and for data and SHERPA (right). The lower plots show the ratio of the distribution in data to in the simulation. The red lines indicate a factor two up and down.

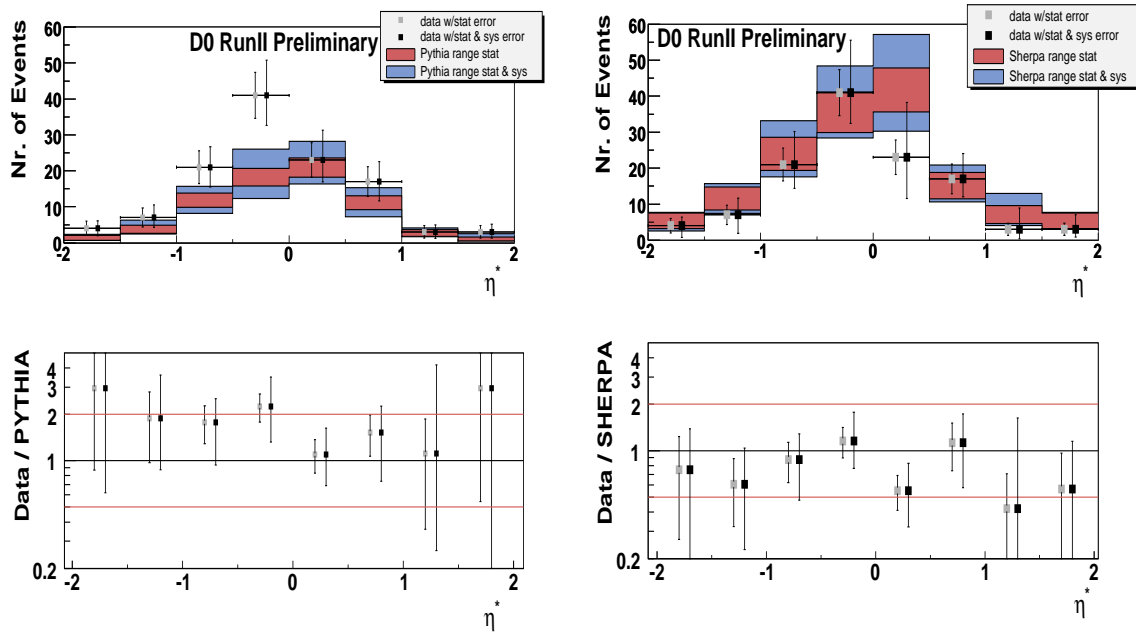


Figure 7.14: Events containing three jets were selected if the the difference in pseudorapidity between the two leading jets exceeds 2.0. The pseudorapidity of the third jet is shown in the coordinate system where zero corresponds to the average pseudorapidity of the two leading jets, for data and PYTHIA (left), and for data and SHERPA (right). The lower plots show the ratio of the distribution in data to in the simulation. The red lines indicate a factor two up and down.

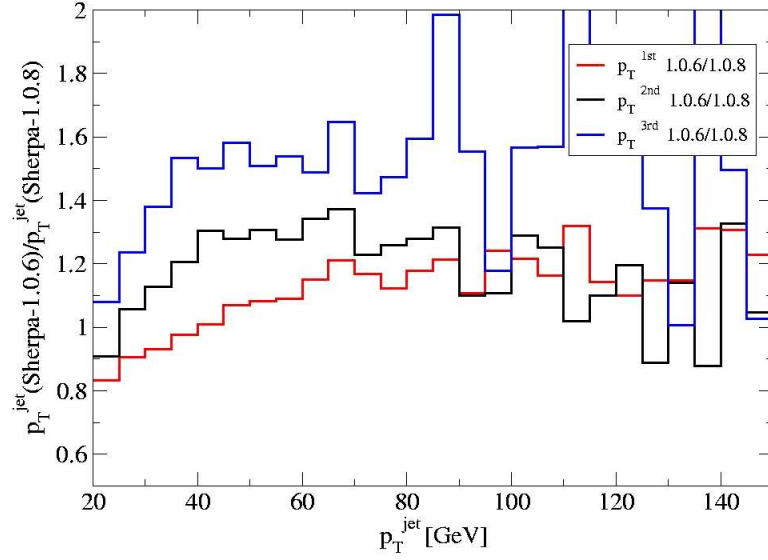


Figure 7.15: Ratios of the three leading p_T^{jet} spectra in $Z/\gamma^*(\rightarrow e^+e^-)+\text{jets}$ production as predicted by SHERPA v1.0.6 to those predicted by SHERPA v1.0.8 [104].

Chapter 8

Correcting Data to Particle Level

The reconstructed spectrum of an observable in data, like p_T^{jet} in $Z + 1\text{-jet}$ events, depends on the performance of the detector. One way of comparing the spectrum reconstructed in data with different theory models is to pass the prediction of each model through a simulation of the detector, as done in the previous Chapter. This process requires large computer resources and access to computer codes which are normally only available to members of the experimental collaboration. To make the data sets useful for theorists, and for the time after the collaboration has ceased to exist, the reconstructed spectrum of the observable must be corrected for impact of the detector. The spectrum is normally corrected to the *particle level*, defined through the list of stable final-state particles predicted by an event generator [105]. A measurement which is corrected to the particle level can be compared directly with the predictions of current and future event generator models without any knowledge about the detector used for the measurement. Below, an observable reconstructed from the detector signals are referred to as the observable at *detector level*, in contrast to the observable defined at the *particle level*.

The main impact of the detector on the spectrum of an observable is:

- Geometrical acceptance: the detector does not cover the whole 4π region, and events where one or more of the final-state objects pass through an un-instrumented part of the 4π phase space are not identified.
- Efficiencies: not all events are registered by the trigger system of the experiment. For events which do pass the trigger, not all reconstructed objects pass the identification criteria.
- Energy scales: the kinematic properties of a reconstructed object differs from the true particle-level properties due to effects like inactive material, uninstrumented regions, noise and pile-up. Energy-scale corrections are applied to assure that $p_T(\text{particle level}) = p_T(\text{detector level})$ is true on average.
- Detector resolution: after energy-scale corrections the reconstructed energies correspond to the particle-level energies, but only on average. On an object-by-object basis significant differences between the detector-level energy and the particle-level energy of the object might remain, depending on the resolution of the detector.
- Misidentification: through misidentification the sample of signal events can receive a contribution from events arising from background processes. For the rest of this Chapter it is assumed that the reconstructed spectrum has been corrected for this by subtracting the estimated background contribution.

Each of these effects influences the reconstructed detector-level spectrum and must be taken into account when correcting the spectrum to the particle-level. Several different methods for correcting a reconstructed observable in data to the particle-level exist. In this Chapter a short overview over different methods for correcting the reconstructed spectrum of an observable to the particle level is given, and the choice of the method used for the measurements presented in Chapter 9 is justified.

8.1 Ansatz-Function-Based Method

The *Ansatz* method for correcting a reconstructed observable to the particle level has been used in several measurements of the properties of jets in multijet events, for instance the $D\bar{O}$ inclusive p_T^{jet} measurement [106]. The first step is to measure the trigger and identification efficiency of the final-state objects, as a function of the observable. Next, the resolution of the observable is measured. A functional ansatz for the particle-level spectrum of the observable is assumed. The trigger and identification efficiencies are applied to the function before it is smeared according to the resolution. The result is referred to as the *detector-folded* ansatz function. The energy scale of the reconstructed data objects is calibrated, so that a given amount of energy, X GeV, at the reconstructed level equals X GeV at particle level. The detector-folded ansatz function is compared with the reconstructed spectrum of the observable in data. Any discrepancies between the two spectra implying that the ansatz function differs from the true particle-level spectrum of data. The ansatz function is iteratively modified until agreement is observed between the detector-folded ansatz function and data. At this point, the ansatz function is equal to the true particle-level distribution in data, within the uncertainties of the efficiencies, the resolution, the energy scale, and the statistical uncertainty in data.

The ansatz method for correcting data to particle level has the advantage of being highly transparent. It is well suited for the measurement of p_T^{jet} in QCD events where high statistics are available to constraint the shape of the ansatz function and to measure the jet trigger and identification efficiencies as a function of the observable using a tag-and-probe method. However, a problem with the ansatz method is that an assumption about the functional form of the particle-level observable in data is needed.

8.2 Event-Generator-Based Ansatz Method

In this method, the Ansatz function is replaced by the particle-level spectrum predicted by an event generator. A particle-level analysis, corresponding to the detector-level analysis used for data, is applied to the simulated event sample. The particle-level objects are rejected according to the trigger and identification efficiencies measured in data. The particle-level energies and directions are smeared to account for the detector resolutions which are measured in data. The resulting spectrum is referred to as the *detector-folded* simulated spectrum. The simulated particle-level spectrum is iteratively reweighted until agreement is observed between the detector-folded simulated spectrum and the reconstructed spectrum in data. At this point, the simulated particle-level spectrum is equal to the true particle-level spectrum in data within the uncertainties of the efficiencies, the resolution, the energy scale and the statistical uncertainty in data.

The event-generator-based ansatz method can be used whenever the normal ansatz method

can be used. In addition, it is practical when the main detector smearing only affects the observable indirectly: examples are angular measurements like $\Delta\phi(\text{jet}, \text{jet})$ in dijet events. The angular resolution for jets is normally very good, and the main resolution effect enters indirectly through the p_T^{jet} resolution. The normal ansatz method with a one-dimensional function does not work in this case, and one would need a two-dimensional ansatz function in the observable and the variable which is smeared by the detector resolution, and such a multi-dimensional ansatz function is most easily accessible through the particle-level prediction of an event generator.

8.3 Full-Detector-Simulation Method

Another alternative for correcting data to the particle level is to rely on a simulated event sample passed through a full simulation of the detector. For the $Z/\gamma^*(\rightarrow ee)+\text{jets}$ measurements presented in Chapter 9 the electron trigger and identification efficiencies are measured using a tag-and-probe method with $\Delta\phi(e, e)$ required to be close to π to suppress backgrounds arising from multijet events. This cut removes a large fraction of the signal events containing two or more jets, and it is therefore not possible to measure the electron efficiencies as a function of jet-properties like $p_T(\text{second jet})$. Instead, the lepton efficiencies are measured in the full, jet-inclusive Z/γ^* sample. Having corrected the lepton efficiencies in the jet-inclusive sample in simulation to those observed in data, the efficiency as a function of the jet observables can be determined in the simulated sample. The simulated sample is in this way used to extrapolate the lepton efficiencies from the inclusive sample, where they can be measured in data, to the exclusive samples containing one or more jets. This extrapolation makes both the Ansatz method and the event-generator-based ansatz method unsuitable for $Z+\text{jets}$ measurements.

The resolution of the detector leads to deviations in shape between the reconstructed spectrum and the particle-level p_T^{jet} spectra. After calibrating the jet energies the reconstructed p_T of a jet will be distributed approximately according to a Gaussian function whose center is the p_T value of the jet at the particle level, and whose width is equal to the jet energy resolution of the detector. For the steeply falling p_T^{jet} spectrum the net effect of the resolution is a migration of events towards higher values of p_T^{jet} . The size of the migration depends on the shape of the distribution.

8.4 Regularized versus Bin-By-Bin Unfolding

In the simulated event sample a binned *migration* matrix, \mathbf{M}_{ij} , consisting of the event-by-event value of observable at the detector level versus at the particle level can be built. The matrix is normalized so that each column contains the probability that an event belonging in bin j at the particle level is reconstructed in bin i at detector level. If the binned particle-level spectrum is written as a vector, \mathbf{p} , the binned detector-level spectrum \mathbf{d} is given by

$$\mathbf{d} = \mathbf{M} \cdot \mathbf{p}. \quad (8.1)$$

In other words, the true particle-level spectrum is folded with the detector resolution, represented by the migration matrix, yielding the detector-level spectrum. Correcting the detector-level spectrum in data for the impact of the detector resolution means solving equation 8.1

for the particle-level spectrum, giving

$$\mathbf{p} = \mathbf{M}^{-1} \cdot \mathbf{d}. \quad (8.2)$$

In practice, both \mathbf{d} and \mathbf{M} are only known with a limited statistical precision, and the right-hand-side of the equation is highly sensitive to small fluctuations in \mathbf{d} . To illustrate this, two statistically independent toy-MC samples were generated. Each sample contains events consisting of pairs of $p_T(\text{jet, particle level})$ and $p_T(\text{jet, detector level})$. The $p_T(\text{jet, particle level})$ values are drawn from a distribution which corresponds to the particle-level $p_T(\text{leading jet})$ distribution for $Z/\gamma^*(\rightarrow ll)+\text{jet}$ events as predicted by ALPGEN+PYTHIA. Each $p_T(\text{jet, particle level})$ value is smeared with the Gaussian corresponding to the jet energy resolution of the DØ detector to define $p_T(\text{jet, detector level})$. The $p_T(\text{jet, particle level})$ – $p_T(\text{jet, detector level})$ pairs define a migration matrix, as well as particle-level and detector-level p_T^{jet} spectra. Two toy-MC samples were generated based on the same ALPGEN+PYTHIA p_T^{jet} spectrum, with one sample containing a factor of ten more p_T^{jet} -pairs than the other. The high (low) statistics sample is analogous to the simulated (data) sample in a real measurement and is labelled *MC* (*data*) below. The particle-level distribution in data, as estimated from the detector-level data distributions by simply multiplying with the inverted migration matrix from the MC sample contains large fluctuations and negative bin contents, see Fig. 8.1. The direct inversion of the migration matrix is clearly not suitable for correcting data to the particle level, even for the idealized case where both data and simulation share the same true particle-level distributions.

The fluctuations seen in Fig. 8.1 can be removed using a regularized unfolding technique which introduces the requirement that either the unfolded distribution is smooth, or that the ratio of the unfolded distribution to the particle-level distribution in simulation is smooth. More specifically, instead of determining \mathbf{p} using Eq. (8.2), one determines \mathbf{p} yielding the minimal value of

$$\mathbf{d} - \mathbf{M} \cdot \mathbf{p} + \tau \cdot \mathcal{F}(\mathbf{p}). \quad (8.3)$$

Here, $\mathcal{F}(\mathbf{p})$ is an additional regularization term which is small (large) if \mathbf{p} has small (large) fluctuations between neighboring bins. For example, $\mathcal{F}(\mathbf{p})$ can be the second derivative of the distribution. The parameter τ is a scalar number which is referred to as the *regularization parameter*. For $\tau = 0$ the particle-level distribution \mathbf{p} which minimizes Eq. (8.3) is equal to the exact solution of Eq. (8.2). For $\tau \rightarrow \infty$ any linear spectrum, having a vanishing second derivative, will minimize the value of Eq. 8.3. For suitable values of τ the identified \mathbf{p} will have the property that it is both approximately a solution of Eq. (8.2) and reasonably smooth, thereby avoiding the large fluctuations of the exact solution of Eq. (8.2) illustrated in Fig. 8.1. Examples of regularized unfolding algorithms, which also contain techniques for identifying a suitable value of τ , are GURU [107] and RUN [108].

A simpler alternative to regularized unfolding is to use

$$\mathbf{p}_{\text{data}} \approx \frac{\mathbf{p}_{\text{sim}}}{\mathbf{d}_{\text{sim}}} \cdot \mathbf{d}_{\text{data}}. \quad (8.4)$$

This is referred to as bin-by-bin unfolding since it does not take migration between the different bins explicitly into account. As for the ansatz-based methods, the simulated particle-level spectrum can be reweighted until agreement is seen between simulation and data at the detector level, thus assuring that Eq. (8.4) is satisfied.

Before deciding on whether or not to use the full migration matrix method or the simpler bin-by-bin method, one can estimate the sensitivity of the measured particle-level distribution

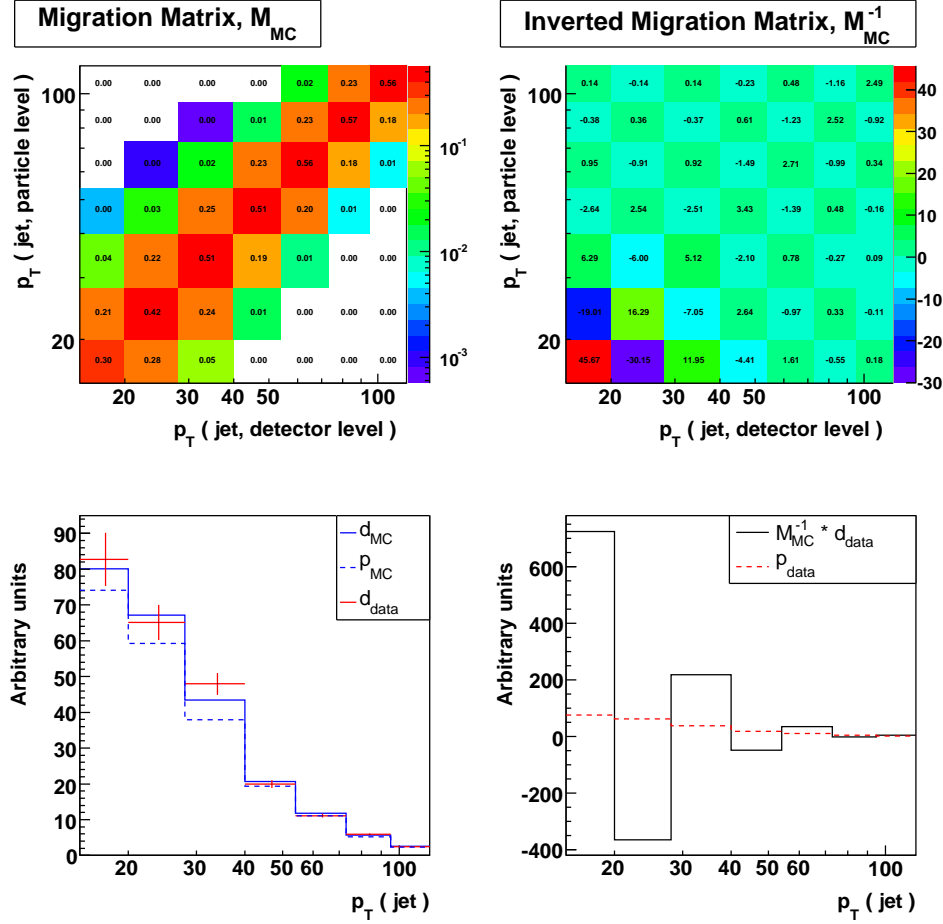


Figure 8.1: The migration-matrix of a high-statistics sample, labelled MC (upper, left). The inverse migration matrix contains large fluctuations between neighboring bins (upper, right). A $data$ sample was generated according to the same particle-level spectrum as the MC sample, but with a factor of ten less events. The particle-level and detector-level MC distributions are shown together with the detector-level data distribution (lower, left). The number of events in the MC sample was normalized to the data sample. The detector-level data distribution is unfolded by multiplying with the inverse of the migration-matrix of the MC samples. The resulting distribution (lower, right) shows large fluctuations and does not give a good estimate of the genuine truth-level distribution in data.

in data to potential disagreements in the shape of the observable between simulation and data. This can be done using a toy-MC study like the one presented in Fig. 8.1, as illustrated below. The particle-level distribution in data is estimated twice: first, by using regularized unfolding taking the full migration matrix into account; and second, by using bin-by-bin unfolding with reweighting of the particle-level spectrum in MC to match data at detector level. Comparisons between the estimated and the true \mathbf{p} gives the accuracy of each method. The accuracy can be studied as a function of the shape differences between the two samples by reweighting one of them at the particle level as a function of the observable.

However, these tests of the accuracy of the unfolding require that a regularized unfolding technique is implemented. An easier test of the need for regularized unfolding is found by correcting from particle to detector level instead of the other way around. Then there are no complications from the inversion of the matrix, and multiplying with M versus with $\frac{\mathbf{p}}{\mathbf{d}}$ can be compared directly without the need of regularization. This was done for the Z +jets analysis presented in Chapter 9 using various toy Monte Carlo studies. The jet energy resolution measured in $D\bar{O}$ Z/γ^* +jet events was used to smear p_T (particle level) values drawn from the particle-level p_T^{jet} spectrum in Z +jet events as predicted by ALPGEN+PYTHIA, thus defining pairs of p_T (particle level) and p_T (detector level) values. This toy-MC sample, labelled A , has the particle-level spectrum \mathbf{p}_A , the detector-level spectrum \mathbf{d}_A , and the migration matrix \mathbf{M}_A . A second sample of p_T pairs, labelled B , was generated using a different spectrum, giving \mathbf{p}_B , \mathbf{d}_B , and \mathbf{M}_B . The discrepancy between $\mathbf{M}_A \cdot \mathbf{p}_B$ or $(\mathbf{d}_A/\mathbf{p}_A) \cdot \mathbf{p}_B$ and the true detector-level distribution \mathbf{d}_A defines the *bias* of the two methods. Three different spectra for the sample B were tested, and the resulting studies are presented in Fig. 8.2. In the first case, B differs from A only by having a less steeply falling slope, corresponding to the ratio $\mathbf{p}_B/\mathbf{p}_A$ increasing from unity at $p_T \approx 20$ GeV to two at $p_T \approx 200$ GeV. The migration-matrix method has a bias at the per-mille level, whereas the bin-by-bin method has a bias of $\sim 2\%$. In the second case, sample B contains a Gaussian peak with 5 GeV width, centered at 150 GeV, which is not included in sample A . The migration-matrix method still has a bias at the per-mille level. The bin-by-bin method fails to take the impact of the resolution on the peak into account, resulting in a bias of up to 60% around 150 GeV. In the third case, a Gaussian peak with a width of 5 GeV was inserted in sample A as well, but centered on 155 GeV instead of 150 GeV.

In conclusion, the bin-by-bin method has an acceptably low bias for smooth distributions like p_T^{jet} in $Z/\gamma^*(\rightarrow e^+e^-)$ +jets production. Since this method is simpler and more transparent than regularized unfolding techniques, the bin-by-bin method is the method of choice for the measurements presented in the next Chapter.

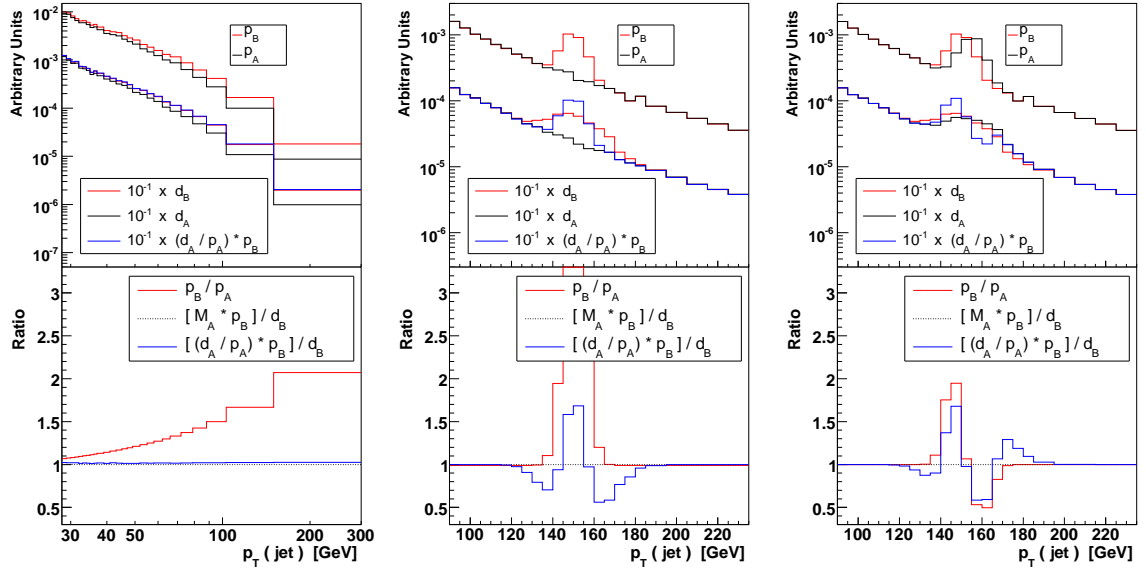


Figure 8.2: Three toy-MC studies illustrating the bias of the migration-matrix method versus the bin-by-bin method. In each toy-MC study, the particle-level spectra of A and B differ in terms of slope (left), in terms of B containing a Gaussian peak (center) and in terms of the Gaussian in sample A being shifted by 5 GeV with respect to sample B (right). In each case, the detector-level spectrum in B (\mathbf{d}_B) is estimated from the particle-level spectrum in B (\mathbf{p}_B) using both the migration matrix of A (M_A) and the bin-by-bin ratio of A ($\mathbf{d}_A / \mathbf{p}_A$). The ratio of the estimated to the true \mathbf{d}_B defines the bias of the correction method. The migration-matrix method has a negligible bias in all three cases. The bin-by-bin method has an acceptably small bias for smoothly falling distributions, but is seen to be unsuitable for more complex spectra.

Chapter 9

Measurements of Differential Cross Sections in $Z/\gamma^* + \text{jets}$ Events

The analysis presented in Chapter 7 compared data with event samples generated using the SHERPA and PYTHIA event generators and a GEANT-based simulation of the performance of the DØ detector. It was argued that the reconstructed distributions should be corrected for the impact of the detector in order to facilitate comparisons with a broader range of theory predictions, and also with different samples from each generator produced using various configuration parameters. In Chapter 8, an overview was given over the various techniques which can be used to correct the spectrum of an observable reconstructed in data for the bias introduced by the detector. It was shown that a method relying on a simulated event sample passed through the GEANT-based detector simulation is most suitable in the case of jet observables in $Z/\gamma^*(\rightarrow e^+e^-) + \text{jets}$ production. In this Chapter, measurements of the differential p_T spectrum of the N^{th} jet¹ in $Z/\gamma^* \rightarrow e^+e^-$ events containing N or more jets are presented, with $N = 1, 2, 3$. The measurements are fully corrected for detector effects and can therefore be compared directly with theory predictions.

9.1 Event Selection

The analysed data sample is described in Sect. 6. Events are required to have been triggered by a single- or di-EM trigger, and physics objects are reconstructed according to the identification criteria described in Chapter 5. An event is assigned as a $Z/\gamma^* \rightarrow e^+e^-$ event (*signal event*) if it contains two electron candidates with opposite-signed electric charge, both satisfying $p_T > 25$ GeV and $|\eta_{\text{det}}| < 1.1$ or $1.5 < |\eta_{\text{det}}| < 2.5$. In order to select events compatible with the decay of a Z/γ^* -boson, the di-electron invariant mass is required to be between 65 and 115 GeV. The reconstructed primary vertex is required to have at least three tracks associated to it and a z coordinate satisfying $|z| < 60$ cm. It is required that the tracks of both electron candidates are compatible with arising from the primary vertex.

Jet candidates are reconstructed using the DØ Run II mid-point algorithm using $R = 0.5$ and the split-and-merge parameter $f = 0.5$ (see Sect. 3.7.1). Jets are required to satisfy $p_T > 20$ GeV and $|\eta_{\text{det}}| < 2.5$, and to have a ΔR -separation to both electrons exceeding 0.4.

A total of 65,759 data events pass the event selection. Of these, 8,452/1,233/167 events have 1/2/3 jets or more, with p_T^{jet} above 20 GeV.

¹Ordered in terms of decreasing values of p_T^{jet} .

9.2 Normalizing the Observables

When using a leading-order event-generator sample to describe data, the sample is usually normalized either to a fixed-order calculation of the cross section, or to the number of events observed in data. Consequently, event generators like SHERPA and ALPGEN+PYTHIA need only to describe differential cross sections relative to the total, jet-inclusive cross section. For the purpose of testing and validating event generators it is therefore sufficient to measure the cross section differential in p_T^{jet} relative to the jet-inclusive cross section, $\frac{1}{\sigma_{Z/\gamma^*}} \times \frac{d\sigma}{dp_T}$, instead of the differential cross section $\frac{d\sigma}{dp_T}$. From an experimental point of view, this has the advantage that any uncertainty which is uncorrelated with the observable in question cancels in the ratio calculation. Two examples are the uncertainty of the integrated luminosity of the data sample, and to some extent of the electron identification efficiency. In addition, the sensitivity to the simulation of higher-order QED corrections in the Z -decay is reduced (see below). The measurement of the jet-inclusive Z/γ^* cross section and comparisons with theory predictions have already been performed in dedicated studies [109, 110], serving as a further motivation for only measuring relative cross sections in this study.

9.3 Backgrounds

There is a certain probability that events not arising from $Z/\gamma^* \rightarrow e^+e^-$ production pass the event selection criteria, giving rise to background events. The sources of background events are classified in two groups: processes containing at least two real electrons², and processes containing one or zero real electrons. The amount of events from the former group is evaluated using simulated event samples described in Sect. 6.2. Each sample was normalized to the measured luminosity using fixed-order cross section predictions at (N)NLO in QCD [55, 34, 111]. The production channels $Z/\gamma^* \rightarrow \tau^+\tau^-$, W , WW and $t\bar{t}$ are taken into account as sources of background, whereas events arising from resonant production of Z +gauge boson $\rightarrow e^+e^- + 2\text{jets}$ are considered as part of the signal. The resulting estimates of the contributions from events containing one or more real electrons to the selected event sample in data are summarized in Tables 9.1-9.4. It is seen that the background contamination from processes containing at least one real electron in the selected data sample is well below 6% in most bins. The background contamination increases with increasing $p_T(\text{jet})$ and jet multiplicity, but it is always smaller than 0.3 times the statistical uncertainty in the data sample. In spite of the small cross section for $t\bar{t}$ production, this channel is the most significant source of background for these measurements. The selected $t\bar{t}$ events are thought to be from the decay mode $t\bar{t} \rightarrow 2e2\nu2b$, meaning that the two leading jets are described by PYTHIA using matrix element calculations which are expected to be relatively accurate. PYTHIA has been found to give a good description of the kinematics of $t\bar{t}$ production in various $D\bar{O}$ studies. Assuming a 10% uncertainty for the $t\bar{t}$ cross section, the uncertainty of the $t\bar{t}$ background prediction is well below 1%, and this is considered to be negligible compared with the main uncertainties of the measurement. The PYTHIA estimate for the $Z/\gamma^* \rightarrow \tau^+\tau^-$ background was validated using ALPGEN+PYTHIA and no significant changes were seen. The estimated amount of background events is subtracted from the data sample, and after this step the data sample is referred to as the *background-subtracted* data sample.

²The term “real electron” is used for prompt electrons only: an electron arising from jet fragmentation is excluded from this definition.

N_{Data}	$\Sigma_{\text{Backgrounds}}$	$\frac{\Sigma_{\text{Backgrounds}}}{N_{\text{Data}}}$	$\frac{\Sigma_{\text{Backgrounds}}}{\sqrt{N_{\text{Data}}}}$	$t\bar{t}$	$Z_{\tau\tau}^{60-130}$	WW	$W_{e\nu}$
65759	114.4	0.00	0.4	11.3	65.3	19.9	17.9

Table 9.1: Background contributions to the jet-inclusive $Z/\gamma^* \rightarrow e^+e^-$ sample.

Binning	N_{Data}	$\Sigma_{\text{Backgrounds}}$	$\frac{\Sigma_{\text{Backgrounds}}}{N_{\text{Data}}}$	$\frac{\Sigma_{\text{Backgrounds}}}{\sqrt{N_{\text{Data}}}}$	$t\bar{t}$	$Z_{\tau\tau}^{60-130}$	WW	$W_{e\nu}$
15 - 20	5782	6.3	0.00	0.1	0.1	3.7	1.8	0.7
20 - 28	3802	6.4	0.00	0.1	0.2	2.7	1.4	2.1
28 - 40	2350	4.6	0.00	0.1	1.0	1.3	0.9	1.4
40 - 54	1081	3.4	0.00	0.1	1.7	1.3	0.4	0.0
54 - 73	639	4.3	0.01	0.2	3.0	0.7	0.7	0.0
73 - 95	322	3.2	0.01	0.2	2.4	0.7	0.1	0.0
95 - 120	145	1.4	0.01	0.1	1.3	0.0	0.1	0.0
120 - 154	72	1.1	0.02	0.1	1.1	0.0	0.0	0.0
154 - 200	35	0.3	0.01	0.1	0.3	0.0	0.0	0.0
200 - 300	6	0.2	0.03	0.1	0.2	0.0	0.0	0.0

Table 9.2: Background contributions to the $p_T(1^{\text{st}} \text{ Jet})$ spectrum (p_T in GeV).

Binning	N_{Data}	$\Sigma_{\text{Backgrounds}}$	$\frac{\Sigma_{\text{Backgrounds}}}{N_{\text{Data}}}$	$\frac{\Sigma_{\text{Backgrounds}}}{\sqrt{N_{\text{Data}}}}$	$t\bar{t}$	$Z_{\tau\tau}^{60-130}$	WW	$W_{e\nu}$
15 - 20	1714	1.4	0.00	0.0	1.1	0.0	0.3	0.0
20 - 28	728	2.6	0.00	0.1	1.5	0.7	0.5	0.0
28 - 40	313	2.9	0.01	0.2	2.3	0.3	0.3	0.0
40 - 54	124	1.8	0.01	0.2	1.7	0.0	0.1	0.0
54 - 73	45	1.2	0.03	0.2	1.2	0.0	0.0	0.0
73 - 200	23	1.3	0.06	0.3	1.3	0.0	0.0	0.0

Table 9.3: Background contributions to the $p_T(2^{\text{nd}} \text{ Jet})$ spectrum (p_T in GeV).

Binning	N_{Data}	$\Sigma_{\text{Backgrounds}}$	$\frac{\Sigma_{\text{Backgrounds}}}{N_{\text{Data}}}$	$\frac{\Sigma_{\text{Backgrounds}}}{\sqrt{N_{\text{Data}}}}$	$t\bar{t}$	$Z_{\tau\tau}^{60-130}$	WW	$W_{e\nu}$
15 - 20	403	1.8	0.00	0.1	1.1	0.3	0.5	0.0
20 - 28	124	1.2	0.01	0.1	0.8	0.3	0.1	0.0
28 - 44	37	0.4	0.01	0.1	0.4	0.0	0.0	0.0
44 - 60	6	0.2	0.03	0.1	0.2	0.0	0.0	0.0

Table 9.4: Background contributions to the $p_T(3^{\text{rd}} \text{ Jet})$ spectrum (p_T in GeV).

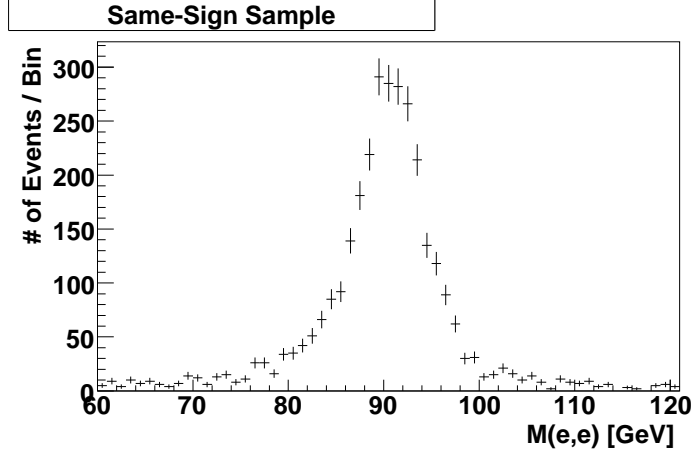


Figure 9.1: Distribution of the di-electron invariant mass in the same-sign sample.

Due to the low probability that a non-electron object is reconstructed as an electron (*fake electron*), a very large sample of simulated multijet and $\gamma + \text{jet}$ events would be needed to get an accurate prediction of the background contribution from these processes (*fake-electron background*). Prohibitively large computer resources would be needed for performing a full detector simulation and the fake-electron background is instead estimated using data events. The signal selection requires two electron candidates with opposite electric charge, defining the opposite-sign (OS) sample. The same event selection, but requiring two electron with the same electric charge defines the same-sign (SS) sample. The charges of two fake electrons are assumed not to be correlated, and consequently the amount of fake-electron events in the SS sample represent an estimate of the amount of fake-electron background events in the OS sample. The di-electron invariant mass spectrum in the jet-inclusive SS sample is given in Fig. 9.1. Due to the non-negligible probability of the charge of a real electron being falsely reconstructed, $Z/\gamma^* \rightarrow e^+e^-$ production also contributes to the SS sample, and indeed a dominant structure compatible with the Z -peak is seen in the SS sample. An upper limit for the fake-electron background contribution can be estimated from the mass ranges corresponding to 20 – 30 GeV above and below M_Z . There are 82 (39) events in the range 61 – 71 GeV (111 – 121 GeV). Assuming these events to be fake-electron events and a linear background shape under the Z -peak, the estimated background in the 65 – 115 GeV mass range is $(82 + 39) * (115 - 65)/20 \approx 300$, corresponding to 0.5% of the OS sample. This is assumed to be an overestimation of the true background since it ignores the contribution of $Z/\gamma^* \rightarrow e^+e^-$ with a falsely reconstructed electron charge to the 61 – 71 GeV and 111 – 121 GeV mass ranges of the SS sample. The estimated fake-electron background in the jet-inclusive data sample is negligible compared with the accuracy aimed for in this measurement and is therefore not taken further into account. The same method is applied to estimate the fake-electron background as a function of p_T^{jet} for the three leading jets, and again no significant contribution was observed (see Fig. 9.2).

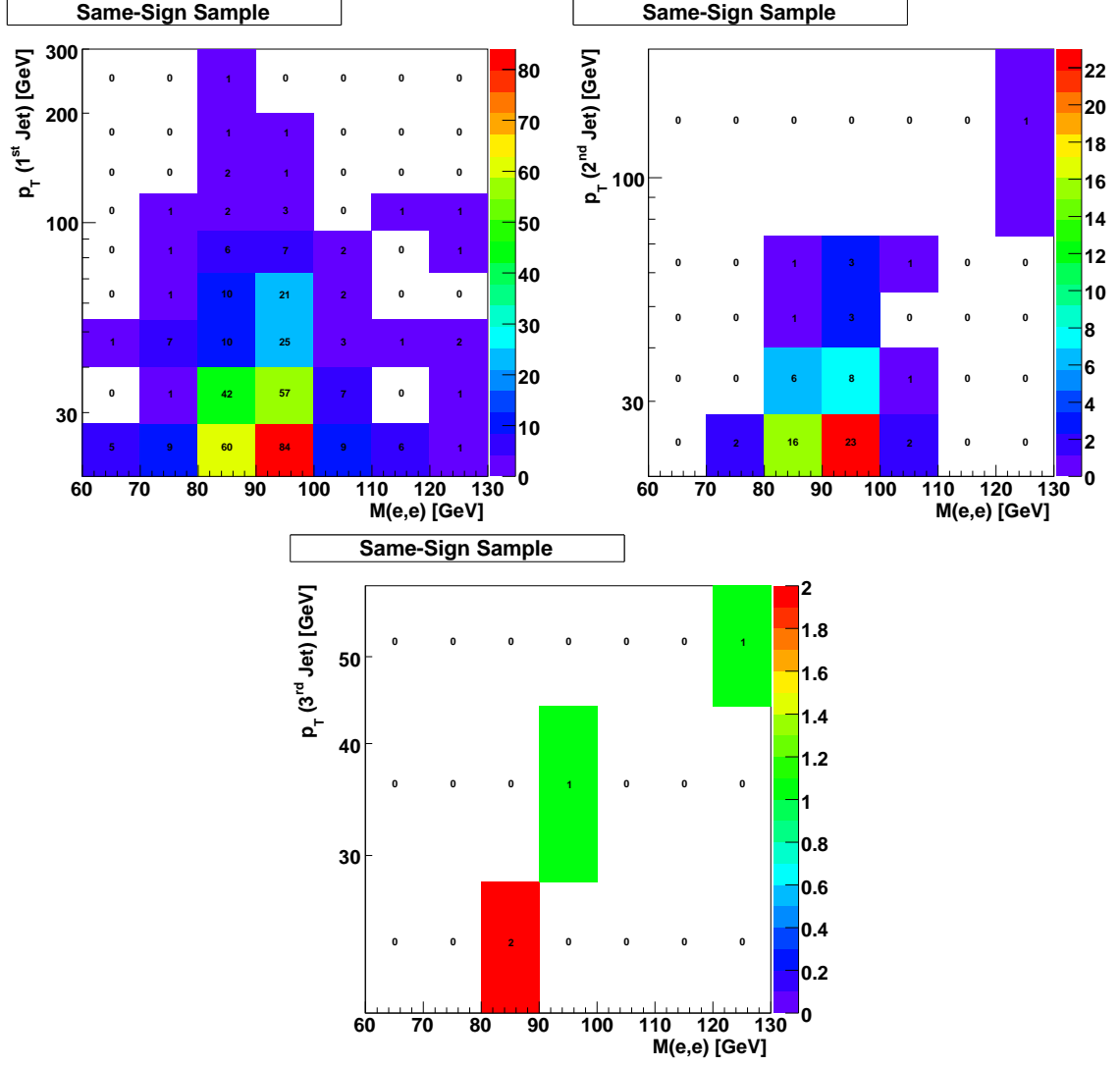


Figure 9.2: Distribution of the di-electron invariant mass vs $p_T(N^{\text{th}} \text{ jet})$ in same-sign events with at least N jets for $N = 1$ (top, left), $N = 2$ (top, right) and $N = 3$ (bottom).

9.4 Correction Data to Particle Level

In this analysis the p_T^{jet} distributions reconstructed in data are corrected for detector effects back to the particle level, using events generated with ALPGEN+PYTHIA (tune A) and passed through the full DØ detector simulation and event reconstruction chain. The same selection as used for reconstructed events in data is applied to the reconstructed events in simulation (referred to as the *detector-level analysis* below). For the simulated event sample an additional analysis is performed using the true particles produced by the event generator (referred to as the *particle-level analysis* below), i.e. using the information which would have been registered by an optimal detector which delivers infinitely precise measurements of the kinematics and flavours of all *stable* incoming or outgoing particles involved the $p\bar{p}$ interaction, with *stable* being defined through a lower limit on the proper life time ($c\tau$), namely $c\tau > 5$ mm. This value is chosen to be well below the inner radius of the beam pipe (18.5 mm) – all later decays being treated by the detector simulation code. This is in agreement with a Les Houches accord [105] specifying that measurements should not rely on model-dependent event-history information.

The event selection in the particle-level analysis is chosen to be analogous to the detector-level selection: $p_{T1} > 25$ GeV, $|y^e| < 1.1$ or $1.5 < |y^e| < 2.5$, and the di-electron system is defined by the two leading electrons requiring opposite-signed electric charge and $65 < M_{ee} < 115$ GeV. Particle-level jets are defined using the same jet algorithm as for reconstructed-level jets, with the list of input objects being all stable particles except the two electrons associated with the Z decay as well as any photon with a ΔR -separation to these two electron of less than 0.2. This latter criterion removes photons which at detector-level are reconstructed as part of the electron object. All jets are required to have a $|y| < 2.5$ and a ΔR -separation to both electrons exceeding 0.4. The final measurements are performed for $p_T^{\text{jet}} > 20$ GeV, but softer jets are used for the study of systematic uncertainties.

For the ideal case that the event generator model perfectly simulates the true $p\bar{p}$ collisions, and that the simulated detector describes the real detector perfectly, the particle-level distribution of an observable in data can be estimated by

$$O_{\text{part}}^{\text{data}} = \frac{O_{\text{part}}^{\text{sim}}}{O_{\text{reco}}^{\text{sim}}} \times O_{\text{reco}}^{\text{data}} = (\mathcal{A} \times \epsilon)_{\text{sim}}^{-1} \times O_{\text{reco}}^{\text{data}}, \quad (9.1)$$

with

$$(\mathcal{A} \times \epsilon)_{\text{sim}} \equiv \frac{O_{\text{reco}}^{\text{sim}}}{O_{\text{part}}^{\text{sim}}} \quad (9.2)$$

where $O_{\text{reco/part}}^{\text{data/sim}}$ is the distribution of the observable in data or simulation at the reconstructed or particle level. The factor $(\mathcal{A} \times \epsilon)_{\text{sim}}$ is referred to as *efficiency times acceptance*. For an ideal detector the factor $(\mathcal{A} \times \epsilon)$ is unity. The acceptance \mathcal{A} for $Z/\gamma^* \rightarrow e^+e^-$ production specifies the fraction of events where both electron pass through an instrumented part of the detector. The efficiency, ϵ , is given by the probability that both electrons are reconstructed given that both enter an instrumented part of the detector. In the present analysis the efficiency and acceptance are not studied separately, but as a product. For steeply falling distributions the detector resolution for the observable in question gives a net migration of events from particle to detector level from bins with a large number to bins with small number of entries, and this effect influences the product $(\mathcal{A} \times \epsilon)$ in each bin. An assumption when correcting data to the particle level is that

$$(\mathcal{A} \times \epsilon)_{\text{data}} = (\mathcal{A} \times \epsilon)_{\text{sim}}. \quad (9.3)$$

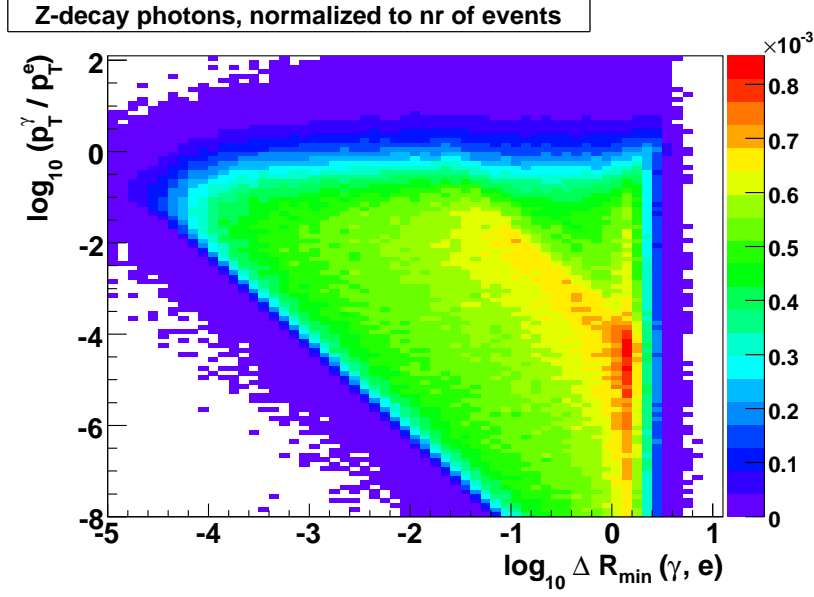


Figure 9.3: For each photon from the Z -decay in the simulated event sample the \log_{10} of the ΔR separation to the nearest electron electrons plotted versus the 10-logarithm of ratio p_T^γ/p_T^e . The distribution is normalized to the total number of events in the sample.

To assure that this is the case, the performance of the simulated detector is corrected to correspond to that of the real detector. Additionally, to make sure that the impact of the migration effects is equal in data and simulation, the shape of the particle-level distribution in simulation is corrected to be equal to that in data. The uncertainty of the measured particle-level distribution in data comes from the uncertainty of the estimated background contributions, and from the uncertainty of the corrections applied to the simulated event sample to assure that Eq. (9.3) is valid.

9.5 Detector-Level versus Particle-Level Electrons

The measurement uses the particle-level analysis to correct the reconstructed data distribution for detector effects. It is therefore a useful check of consistency to establish that the objects which enter the detector-level and particle-level are related to each other in a one-to-one relationship.

One difference between the particle- and detector-level selections is the definition of an electron. At leading order, the electrons arise from the decay $Z/\gamma^* \rightarrow e^+e^-$. When including higher orders correction the possible decays are $Z/\gamma^* \rightarrow e^+e^- + N \cdot \gamma$, with $N = 0, 1, 2, \dots$. In Fig. 9.3 the minimal ΔR separation for each photon (γ) from Z -decay to the two electrons (e) from the Z -decay is plotted versus the ratio p_T^γ/p_T^e for the ALPGEN sample. The total number of events was normalized to data. As expected, it is seen that the photons tend to be soft and/or collinear with respect to an electron. For collinear $e-\gamma$ pairs, the γ frequently carries a large fraction of the total momentum of the pair. At detector level, electrons are based on $R = 0.2$ cones of energy deposits in the EM calorimeter. Reconstructed electron objects

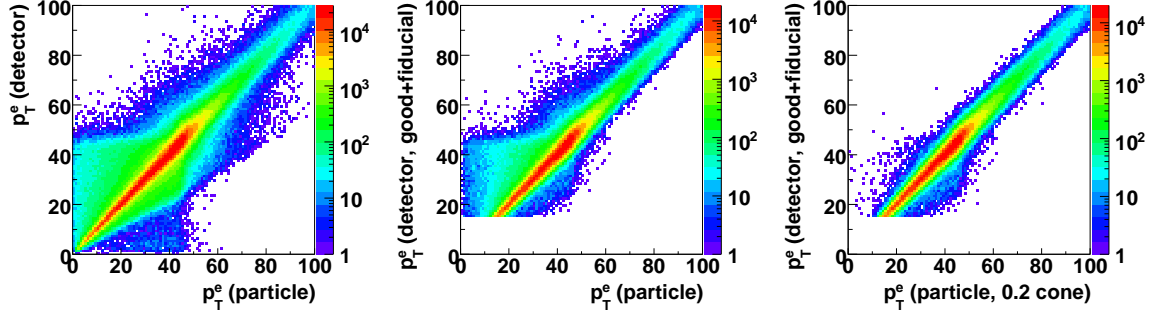


Figure 9.4: The p_T correlation of pairs of one particle-level and one detector-level electron using the matching criterion $\Delta R < 0.2$. The correlation is defined once using particle-level electrons and all reconstructed electron candidates (left); once requiring reconstructed electrons to be fiducial and to pass the ID criteria (center); and redefining the particle-level object as consisting of the sum of electrons and any photon in a cone $R = 0.2$ around it (right).

correspond to the sum of electrons and photons within the $R = 0.2$ cone. At particle level, in contrast, an electron does not take nearby photons into account, regardless of proximity in R , and this leads to off-diagonal elements in the migration matrix between particle-level and detector-level p_T^e (Fig 9.4). Off-diagonal elements also arise from reconstructed close to ϕ -cracks in the calorimeter. About 5% of the entries in the migration matrix have $p_T^{\text{reco}} > 1.2 \times p_T^{\text{particle}}$ or the other way around. A diagonal migration matrix can be recovered by requiring reconstructed electrons to be fiducial, and by defining particle-level electron objects as the sum of each electron and any photon in an $R = 0.2$ cone around it. However, since the Z decay mode and the ϕ_{det} of the electron are not expected to be correlated with the p_T^{jet} spectra, the effect of off-diagonal entries cancels in the normalized observable $\frac{1}{\sigma_{Z/\gamma^*}} \times \frac{d\sigma}{dp_T}$. To gain statistics reconstructed electrons are not required to be fiducial, and to simplify the particle-level definition the electron object is not redefined to recover close-by photons.

9.6 Consistency checks of the Jet Definition

The default jet algorithm is the DØ Run II Mid-Point algorithm using a cone radius of $R = 0.5$ and split-and-merge parameter $f = 0.5$. To validate that the clustered jets correspond to a collimated spray of particles, and that $R = 0.5$ is a suitable choice the measured p_T^{jet} range, the p_T density of particles in particle-level jets in simulation were studied as a function of p_T^{jet} , see Fig. 9.5. Above a p_T^{jet} of 30 GeV the energy density has its maximum at a separation of $\Delta R = 0$, in agreement with a naive association of each jet with a single, high- p_T parton in a leading-order picture. At lower values of p_T^{jet} this picture is clearly invalid, and the maximal energy density is found at $0.1 < \Delta R < 0.2$ for $10 < p_T^{\text{jet}} < 15$ GeV. In addition, at low p_T there is a significant tail towards large values of ΔR , and for $10 < p_T^{\text{jet}} < 15$ GeV $\sim 15\%$ of the jet p_T is carried by particles which do not belong to a naive $R = 0.5$ cone around the jet axis. This can happen due to the split-and-merge prescription of jet algorithm and indicates that a larger cone size might be more suited for probing the p_T^{jet} region below 20 GeV.

As a cross-check that the particle-level jet definition is suitable for probing QCD activity

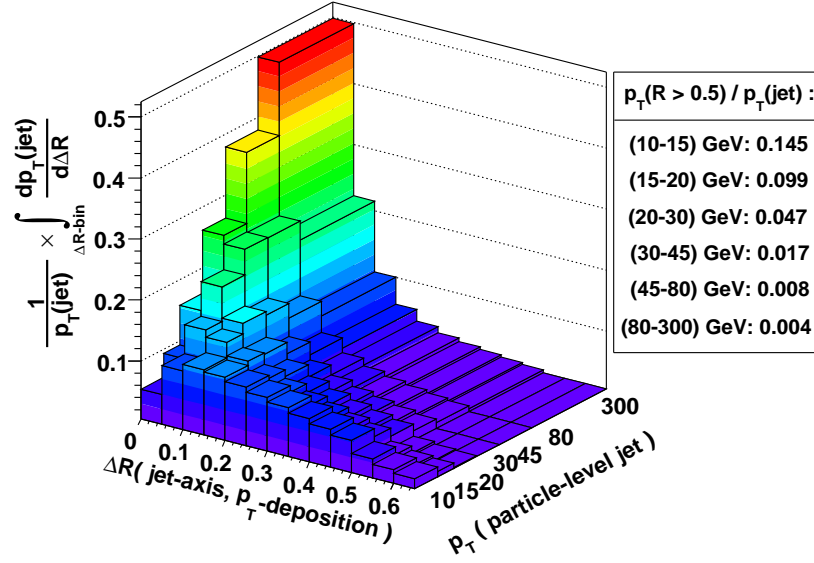


Figure 9.5: The radial momentum density of DØ Run II Mid-Point $R = 0.5$ jets at the particle level, as a function of p_T^{jet} . The average fraction of the jet momentum located outside of the naive 0.5 cone, in each p_T bin, is given in the legend.

the particle content in particle-level jets was studied. For each jet, the list of contributing particles was split in classes of particle flavour. The p_T of the 4-momentum vector sum for each flavour, relative to p_T^{jet} , is shown in Fig. 9.6. Each distribution is normalized to the mean p_T fraction of jets arising from the particle flavour in question. As expected, in most jets the majority of the p_T is carried by mesons, followed by baryons. The photon distribution is peaked both at 0 and 1. The peak at 1 has been traced back to jets consisting mainly of one high- p_T photon from the decay of the Z boson which was radiated at a large angle with respect to both electrons. These particle-level jets do not probe QCD and are not likely to be reconstructed as a detector level jet due to their large electromagnetic energy fraction. This fact introduces sensitivity of the measured p_T^{jet} spectra on the simulation of photons stemming from Z decay. Since on average photons carry only 4% of the jet p_T , and since the simulation of higher-order QED corrections to the Z decays have been tuned to reproduce precision measurements from LEP, this effect is expected to be accurately modeled.

9.7 Choice of binning

The correlation between $p_T(\text{leading jet})$ at detector and particle level is given in Fig. 9.7 (left). The detector resolution of an observable (σ) constrains the minimal size of the chosen bins. The *purity* (*efficiency*) of a bin is defined as the fraction of events which contribute to the bin at detector level (particle level) which contribute to the same bin at particle level (detector level). If $w \ll \sigma$ the purity and efficiency of each bin is low and consequently the number of entries in a given bin at detector level contains relatively little information about the number of entries in the same bin at the particle level. As an example, the purity and efficiency for

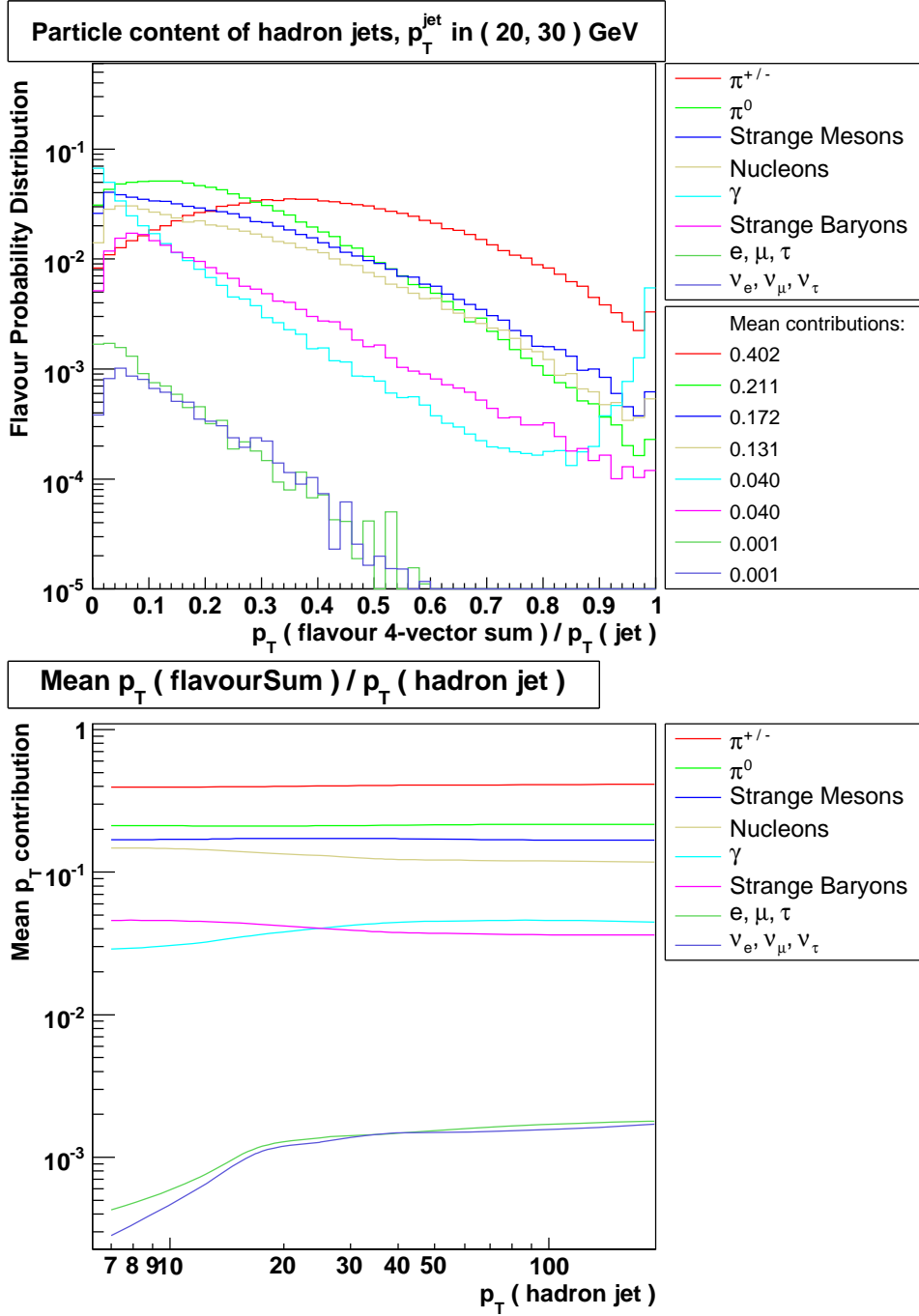


Figure 9.6: The distribution of the p_T fraction carried by each particle flavour for particle-level jets satisfying $20 < p_T^{\text{jet}} < 30$ GeV (top). Each distribution is normalized to the mean p_T fraction of jets arising from the particle flavour in question. As seen in the lower plot, the mean p_T carried by each particle flavour is fairly constant as a function of p_T^{jet} .

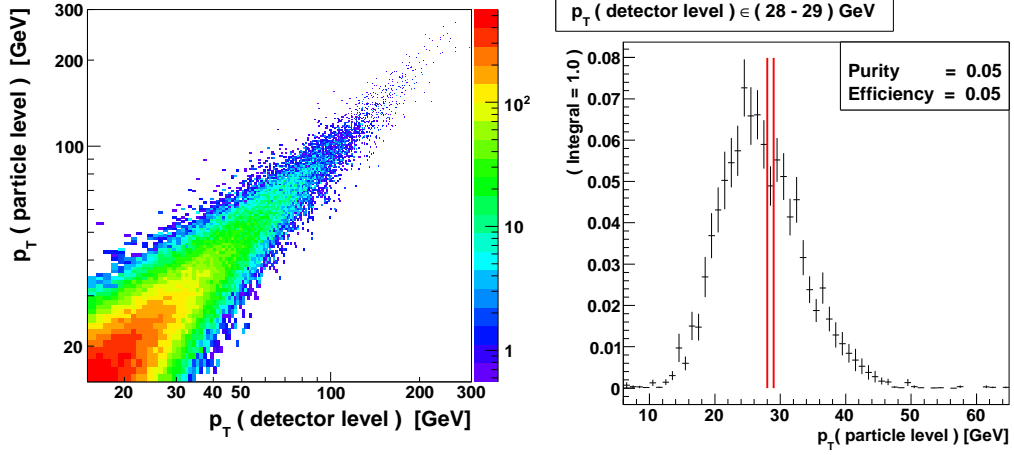


Figure 9.7: The correlation between detector-level and particle-level p_T^{jet} for the leading jet with 1 GeV binning (left). The particle-level p_T^{jet} distribution for jets with detector-level p_T^{jet} in the bins 28 – 29 GeV (right), corresponding to a purity and efficiency of 5%. The range corresponding to the detector-level bin is indicated by vertical, red lines.

the 28 – 29 GeV bin is only 5% as seen in Fig. 9.7 (right). In this case neighbouring bins are highly correlated and this complicates the interpretation of the measurement. To reduce the correlations between different bins a minimal efficiency and purity for each bin can be required. Obviously, the bin width should also not be chosen to be too coarse since this decreases the sensitivity of the measurements to the details of the spectrum.

The binning of the present measurements were chosen such that the efficiency and purities of each bin is in the range 50 – 60%. The correlation between $p_T(\text{leading jet})$ at detector and particle level with the chosen binning is given in Fig. 9.8, and as an example the purity and efficiency of the bin 28 – 40 GeV is seen to be roughly 55%.

9.8 Distinguishing Detector and Generator Effects

In Fig. 9.9 a comparison between the p_T spectrum of the leading jet in data and simulation is given. The simulated event sample is normalized to the measured luminosity of the data sample using a NNLO prediction for the $Z/\gamma^* \rightarrow e^+e^-$ cross section times branching ratio $\sigma_{Z/\gamma^*} \times \text{BR}$ of 241.6 ± 8.7 pb [112, 113]. Above 30 GeV simulation describes the shape of data, but the normalization is $\sim 10\%$ higher. From this one could conclude that the particle-level distribution in simulation, scaled down by $\sim 10\%$, is a good estimate for the particle-level distribution in data for the p_T of the leading jet above $p_T = 30$ GeV. However, as seen below the electron ID efficiency in simulation is $\sim 10\%$ higher than in data, and the ALPGEN+PYTHIA sample contains $\sim 20\%$ less jets than in data. The former discrepancy affects the detector-level p_T^{jet} spectrum only, whereas the latter affects both the particle-level and detector-level spectra in approximately the same way. This illustrates that the agreement between data and simulation at the detector level does not imply that the particle-level distribution in simulation is a good estimate for the particle-level distribution in data.

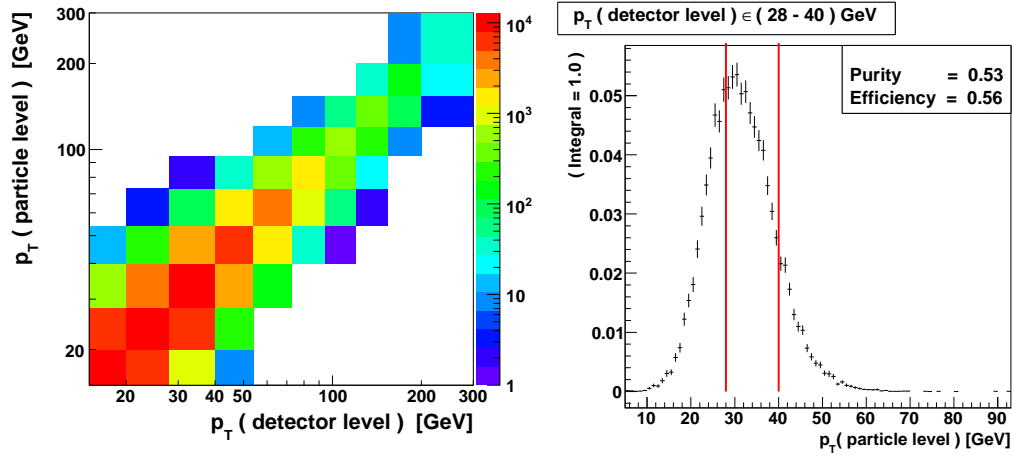


Figure 9.8: The same distributions as shown in Fig. 9.7, but with the choice of binning used for the measurement of $p_T(\text{leading jet})$.

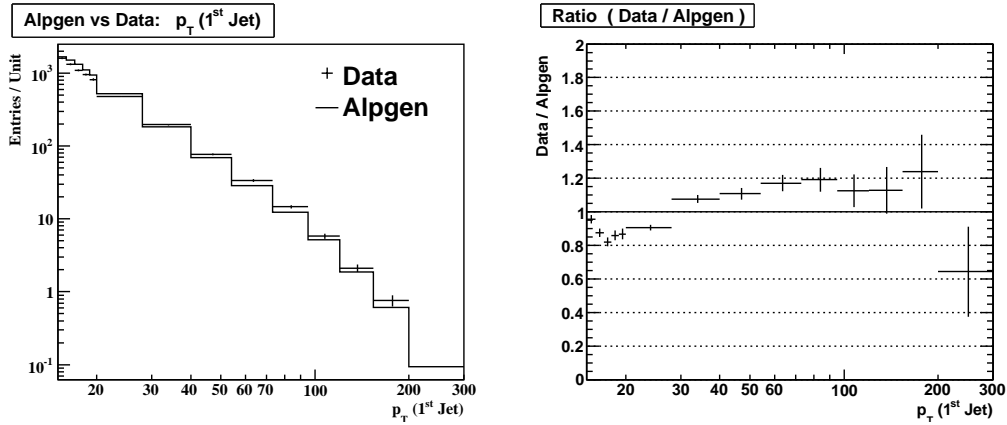


Figure 9.9: The p_T spectrum of the leading jet in data and ALPGEN+PYTHIA.

For a correct estimate for data at particle level, the simulated event sample must be corrected in two separate steps. First, corrections are applied to compensate for any relevant differences between the simulated and the real experimental setups, including both the beam conditions and the detector performance. Next, it is assumed that any remaining differences between data and simulation must arise from a lack of accuracy in the event-generator model, and the simulated event sample is reweighted as a function of particle-level quantities until agreement is seen at the detector level. At this point, the particle-level distribution in simulation is approximately equal to the particle-level distribution data, and $(\mathcal{A} \times \epsilon)_{\text{data}} = (\mathcal{A} \times \epsilon)_{\text{sim}}$ is valid within the uncertainties of the applied corrections. In the following, each correction step is discussed separately. In particular it is studied to which extent the necessary corrections are a function of p_T^{jet} , or whether correction factors derived in the jet-inclusive sample can be applied.

9.9 Collider-specific corrections

The output signals from the simulated detector are added to the corresponding information from zero bias events registered with the real DØ detector to account for effects like noise and pile-up. The energy density in zero bias events increases with increasing average number of $p\bar{p}$ interactions per bunch crossing, and therefore with the Tevatron instantaneous luminosity.

As seen in Fig. 9.10, the instantaneous luminosity profile of the zero bias sample which was added to the simulated events do not accurately describe the profile of the analyzed data sample. The simulated event sample was reweighted as a function of instantaneous luminosity to compensate for this difference.

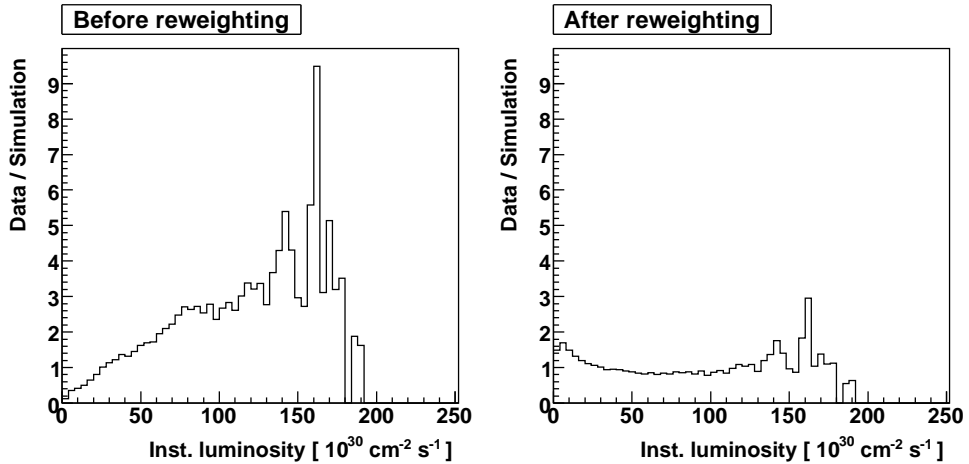


Figure 9.10: The zero bias events overlaid on the simulated events have a different instantaneous luminosity profile than the data set. The ratio of the instantaneous luminosity in data and simulation is shown before (left) and after (right) applying the reweighting procedure.

The distribution of the z coordinate of the primary interaction vertex in simulation is taken to be a Gaussian with a width σ of 25 cm. This does not reflect the corresponding distribution in data where a larger fraction of events have large values of $|z|$ than in simulation. This

affects for instance the identification efficiency of electrons since the track reconstruction efficiency decreases with increasing $|z|$. The simulated event sample is reweighted [114] to reproduce the z distribution measured in data which is shown in Fig. 9.11.

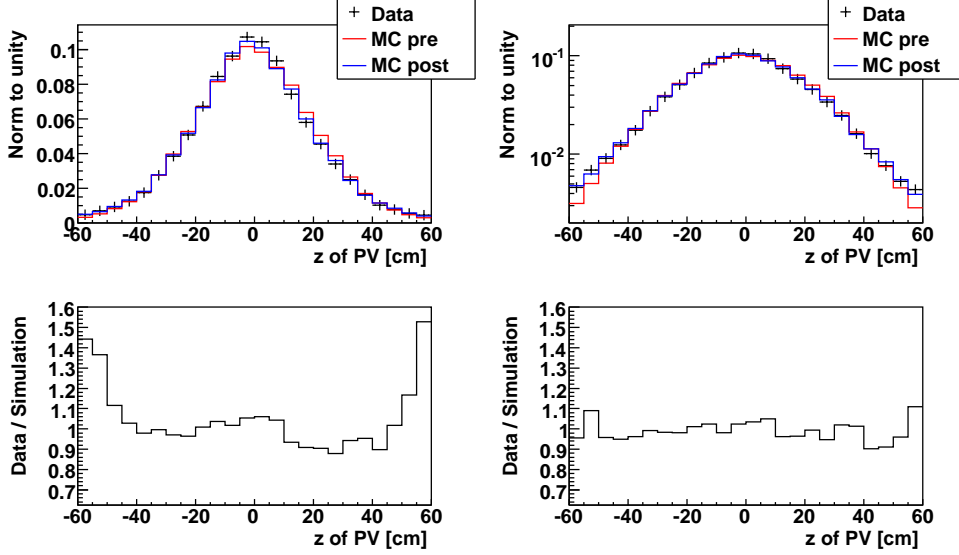


Figure 9.11: The z coordinate shown in data and simulation before (labeled *MC pre*) and after (labeled *MC post*) applying the reweighting for linear scale (upper left) and log scale (upper right). The ratio of data to simulation before and after applying the reweighting are shown on the lower left and lower right, respectively.

9.10 Detector-specific Corrections related to Electrons

The electromagnetic energy scale in data and simulation is defined by requiring that the reconstructed Z mass corresponds to the world-average value of $M_Z = 91.1876 \pm 0.0021$ GeV [1]. Before the corrections the energy resolution is higher in simulation than in data, reflected in a narrower Z mass peak spectrum. To compensate for this deviation, the energies of simulated electrons are smeared with a Gaussian function. The Z mass peak spectra in data and simulation before the last correction step is shown in Fig. 9.12, illustrating the effect of both the energy scale shift and the energy smearing performed to simulation.

The event selection requires data events selected by a single- or di-EM trigger. From earlier measurements, e.g. Ref. [115], it is known that the $D\phi$ trigger efficiency for Z/γ^* events with selection criteria used in the present analysis exceeds 99%. The trigger system is not simulated, meaning that the trigger requirement is applied only in the analysis of the data sample. No trigger efficiency correction is applied to the simulated event sample since the efficiency in data is compatible with 100% with an uncertainty of less than 1%. Additionally, due to the normalization of the jet observables to the inclusive $Z/\gamma^* \rightarrow e^+e^-$ cross section, the effect of a trigger inefficiency would cancel as long as it is not correlated with the observable. The correlation between the trigger efficiency and the jet activity is studied in Sect. 9.13.1.

The probability for identifying electrons is higher in simulation and in data. To compensate for this, reconstructed electrons in simulation are rejected with the probability given by the

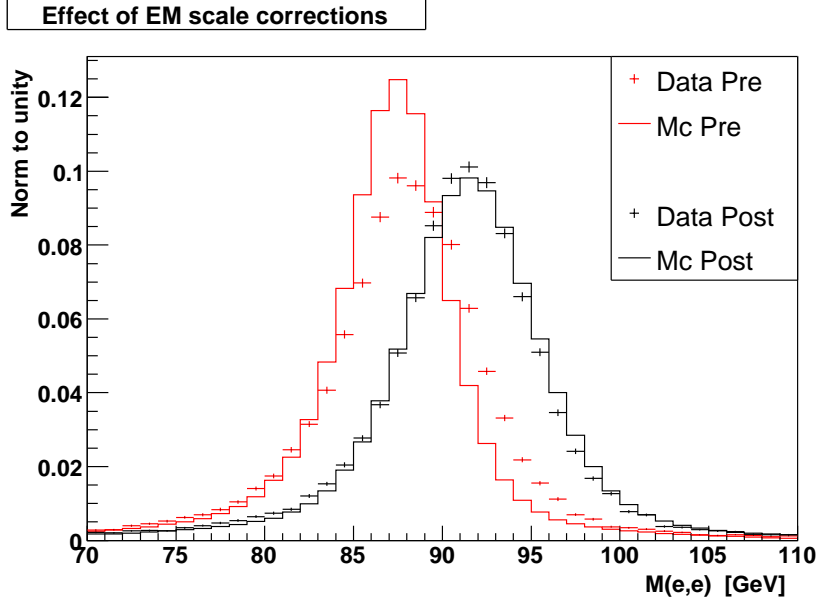


Figure 9.12: The dielectron invariant mass in data and simulation before and after the final EM energy scale correction.

ratio of the ID efficiency in data to that in simulation. This was done using jet-inclusive efficiencies measured 2-dimensionally in bins of η_{det} and ϕ_{det} provided by [88]. The correlation between the electron ID efficiency and the jet activity is studied in Sect. 9.13.2.

9.11 Measuring the Inclusive $Z/\gamma^* \rightarrow e^+e^-$ Cross Section

After applying the corrections described above, the simulated and real detectors have the same performance for identifying and measuring electrons in jet-inclusive events. Consequently, $(\mathcal{A} \times \epsilon)_{\text{data}} = (\mathcal{A} \times \epsilon)_{\text{sim}}$ for the $Z/\gamma^* \rightarrow e^+e^-$ cross section (σ_{incl}). During the data-taking period, the number of signal events is $N_{\text{all}}^{\text{data}}$, according to

$$N_{\text{all}}^{\text{data}} = (\sigma_{\text{incl}} \times \int \mathcal{L} dt). \quad (9.4)$$

Applying the event reconstruction and selection to the data set $N_{\text{reco}}^{\text{data}}$ events were selected, with

$$N_{\text{reco}}^{\text{data}} = (\mathcal{A} \times \epsilon)_{\text{data}} \times N_{\text{all}}^{\text{data}} + N_{\text{BG}}, \quad (9.5)$$

where N_{BG} is the amount of background events. The integrated luminosity of the data sample was determined in Sect. 6.1.2 to be 1073 pb^{-1} with an uncertainty of $\pm 6.1\%$. In Sect. 9.3 the number of background events in the inclusive data sample was estimated to be 114. The $(\mathcal{A} \times \epsilon)$ in simulation is 0.527 using particle selection criteria analogous to in the detector-level analysis, namely $p_{T1} > 25 \text{ GeV}$, $|y^e| < 1.1$ or $1.5 < |y^e| < 2.5$ and $65 < M_{ee} < 115 \text{ GeV}$. During the data taking period, the calorimeter was inactive for $\sim 2.9\%$ of the registered bunch crossings [116]. This effect is not modeled in the simulated event sample, and therefore

$(\mathcal{A} \times \epsilon)_{\text{data}} = (\mathcal{A} \times \epsilon)_{\text{sim}} \times \epsilon^{\text{calo}}$, with $\epsilon^{\text{calo}} = 0.971$. Assuming a conservative 5% uncertainty for the jet-inclusive electron ID efficiency corrections,

$$\begin{aligned}\sigma_{\text{incl}} &= \frac{1}{\int \mathcal{L} dt} \times \frac{N_{\text{Reco}}^{\text{Data}} - N_{\text{BG}}}{(\mathcal{A} \times \epsilon)_{\text{sim}} \times \epsilon_{\text{calo}}} \\ &= \frac{1}{1073 \text{ pb}^{-1}} \times \frac{65759 - 114}{0.527 \times 0.971} \\ &= 119.6 \pm 0.5(\text{stat}) \pm 6.0(\text{sys}) \pm 7.3(\text{lumi}) \text{ pb},\end{aligned}\tag{9.6}$$

with the statistical uncertainty indicated labeled (*stat*), the systematic uncertainty labeled (*sys*) and the uncertainty arising from the integrated luminosity of the data sample labeled (*lumi*).

Alternatively, the measurement can be performed for the particle-level selection consisting of $65 < M_{ee} < 115$ GeV only, corresponding to extrapolating the measurement from the single-electron phase space registered at detector level to the full single-electron phase space. In this case the $(\mathcal{A} \times \epsilon)$ is 0.233, giving

$$\sigma_{\text{incl}} = 270 \pm 1(\text{stat}) \pm 13(\text{sys}) \pm 16(\text{lumi}) \text{ pb}.\tag{9.7}$$

The measured cross sections are used for normalizing the jet measurements as described in Sect. 9.2.

As a consistency check of the present analysis, the inclusive cross section measurement is compared with the result of a dedicated and more detailed $D\bar{O}$ measurement of the Z cross section [117]. That measurement uses $70 < M_{ee} < 110$ GeV, and for this particle-level selection the $(\mathcal{A} \times \epsilon)$ of the present measurement is 0.238, giving

$$\sigma_{\text{incl}} = 265 \pm 1(\text{stat}) \pm 13(\text{sys}) \pm 16(\text{lumi}) \text{ pb}.\tag{9.8}$$

The two measurements are in reasonable agreement as seen in Tab. 9.5. Since the two analysis use overlapping data-sets and the same luminosity calculation, the statistical uncertainties and the luminosity uncertainties are fully correlated. However, this is not the case for the electron efficiency uncertainties due to different electron identification criterias being used. Since the inclusive cross section is only used for normalization purposes in the present study, i.e. any effect not correlated with jet observables cancels, the $\sim 2\%$ deviation between the two measurements seen in Tab. 9.5 has not been investigated any further.

	$\sigma_{\gamma/Z \rightarrow ee}, M_Z \in [70 - 110] \text{ GeV}$
Present analysis	$265 \pm 1(\text{stat}) \pm 13(\text{sys}) \pm 16(\text{lumi}) \text{ pb}$
Ref. [117]	$259.3 \pm 1.6(\text{stat}) \pm 5.0(\text{sys}) \pm 15.8(\text{lumi}) \text{ pb}$

Table 9.5: The $\gamma/Z \rightarrow ee$ cross section for $70 < M_{ee} < 110$ GeV as measured in the present analysis compared with the result from Ref. [117].

9.12 Detector-specific Corrections related to Jets

So far, the simulated detector performance for electrons has been corrected to correspond to that of the real $D\bar{O}$ detector. The jet-inclusive analysis has been measured and validated against a dedicated cross section measurement. Before measuring jet-related observables, the same correction steps must be repeated for the simulated detector performance for jets.

9.12.1 Jet Properties

Reconstructed jet energies are scaled with jet energy scale corrections derived using γ +jet and dijet events (see Sect. 5.5.3). The jet energy scale corrections guarantee that jets in γ +jet events have, on average, identical reconstructed energy and particle-level energy. To account for the difference in jet energy scale between γ +jet events and Z/γ^* +jet events, an additional set of energy corrections were derived for exclusive, back-to-back $Z + 1$ -jet events using the SSR method (see Sect. 5.5.4). The SSR method also smears jet energies in simulation to adjust the jet energy resolution to agree with data.

It should be noted that the measurement only depends on the jet energy scale being the same in data and simulation, not on this energy scale being correct in the sense that the p_T value of a jet is on average the same on the detector level as on the particle level. The jet reconstruction and identification efficiency are measured using a tag-and-probe method in dijet and γ +jet events (see Sect. 5.5.2). In data, the reconstruction efficiency is measured to be $\sim 100\%$ above ~ 40 GeV, while being significantly lower for smaller p_T^{jet} values due to the jet reconstruction threshold (see Fig. 5.4). The jet ID efficiency is $95 - 100\%$ for $20 < p_T^{\text{jet}} < 300$ GeV. The product of the reconstruction and ID efficiencies in simulation is found to be $1 - 2\%$ below that in data, and identified jet candidates are rejected based a probability equal to the ratio of the efficiency in data to that in simulation, binned in η_{det} and p_T^{jet} . An exception is the reconstruction efficiency below ~ 40 GeV where deviations of up to 10% are seen (see Fig. 5.4), arising from the interplay between difference in jet energy response in data and simulation and the jet reconstruction threshold. These latter discrepancies are taken into account using the SSR method.

Topology Dependency of the Jet Energy Scale

The present analysis relies on the jet energy scale and jet energy resolution in simulation and data being the same not for the back-to-back $Z + 1$ -jet sample where the corrections applied to simulation using the SSR method was derived (see Sect. 5.5.4), but for every bin of the measurements of the p_T^{jet} spectra, within uncertainties. To verify that this is the case, parts of the SSR study was repeated for different event selections, starting with the back-to-back exclusive 1-jet selection:

- (1) $\Delta\phi(Z, \text{leading jet}) > 3.0$; $N_{\text{jet}} = 1$;

then, loosening the back-to-back cut,

- (2) no $\Delta\phi(Z, \text{leading jet})$ requirement; $N_{\text{jet}} = 1$;

next, loosening also the veto against sub-leading jets, arriving at the event selections used in the present measurements,

- (3) no $\Delta\phi(Z, \text{leading jet})$ requirement; $N_{\text{jet}} \geq 1$;
- (4) no $\Delta\phi(Z, \text{leading jet})$ requirement; $N_{\text{jet}} \geq 2$;
- (5) no $\Delta\phi(Z, \text{leading jet})$ requirement; $N_{\text{jet}} \geq 3$.

In the original SSR study the kinematic balance in the event is probed through the distribution of

$$\Delta S = \frac{p_T^{\text{jet}} - p_T^Z}{p_T^{\text{jet}}} \quad (9.9)$$

in bins of p_T^Z . The distribution of ΔS using selection (1) is given in Fig 9.13 for $25 < p_T^Z < 30$ GeV and $60 < p_T^Z < 70$ GeV. For the latter p_T^Z bin the distribution corresponds approximately to a Gaussian distribution centered at zero, as expected if $p_T^Z \approx p_T^{\text{jet}}$ with a p_T imbalance arising from the jet energy resolution. For low p_T^Z bins like $20 - 30$ GeV the distribution is asymmetric due to the bias arising from the $p_T^{\text{jet}} > 20$ GeV requirement. For large p_T^Z the RMS of the ΔS distribution can be identified with the jet energy resolution, and the mean with the jet energy scale. For small p_T^Z values, below ~ 50 GeV, the bias from the lower p_T^{jet} selection means that both the RMS and the mean contains information from both the jet energy resolution and scale, and one would need to fit a turn-on convoluted with a Gaussian to extract the two quantities separately. For the present study it is assumed that agreement between data and simulation for both the RMS and mean implies agreement for both jet energy resolution and scale.

The RMS and mean of the ΔS distribution ($\langle \Delta S \rangle$) extracted in bins of p_T^Z for both in data and simulation are given in Fig. 9.14, using selection (1). The jet energy scale and resolution corrections have already been applied to the simulated sample, so data and simulation should agree per construction. The structure seen in the RMS spectrum is an artifact of the algorithm used to prevent isolated, outlying entries in the data histogram from affecting the RMS. The algorithm affects the RMS value found in different bins, but not the relative agreement between data and simulation in each single bin.

To facilitate the removal of the $\Delta\phi(Z, \text{jet})$ criterion and of the veto against sub-leading jets, ΔS is redefined as

$$\Delta S = \frac{P_Z(\sum \text{jets}) - p_T^Z}{p_T^Z}, \quad (9.10)$$

where $P_Z(\sum \text{jets})$ is the transverse component of the 4-momentum vector sum of all jets in the event, projected onto the axis in the transverse plane defined by p_T^Z . The RMS and mean of this modified ΔS distribution are extracted in bins of p_T^Z for event selection (2), and the results are given in Fig. 9.15. Agreement is seen within uncertainties. Finally, the veto against sub-leading jets is removed, and the RMS and mean is plotted for events with at least 1 jet (Fig. 9.16) and 2 jets (Fig. 9.17), corresponding to selection (3)-(4) above. For events with 3 or more jets (selection (5)) the limited statistics in data allows only for one p_T^Z bin, and the corresponding ΔS distribution is given in Fig. 9.18. In all cases, data and simulation is found to agree within the uncertainties.

In conclusion, the jet energy scale and resolution corrections applied to simulation are valid within the uncertainties for events with at least $N = 1, 2, 3$ jets without any requirement on the event topology, i.e. for the samples uses to measure $p_T(N^{\text{th}} \text{ jet})$.

9.13 Jet Dependency of Electron Corrections

The electron trigger and ID criteria include the isolation, electromagnetic energy fraction and calorimeter shower shape of the electron objects, hadrons from QCD activity with a small ΔR separation to an electron can lead to the electron failing the selection criteria of the analysis. In case the correlations between the jet activity and selection efficiency is not described by simulation this will occur in a different fraction of the events in data and simulation, and consequently the assumption that $(\mathcal{A} \times \epsilon)_{\text{data}} = (\mathcal{A} \times \epsilon)_{\text{sim}}$ is wrong and the measurement is biased. To account for this, the electron efficiency corrections would have to be measured and applied as a function of the jet activity in the event.

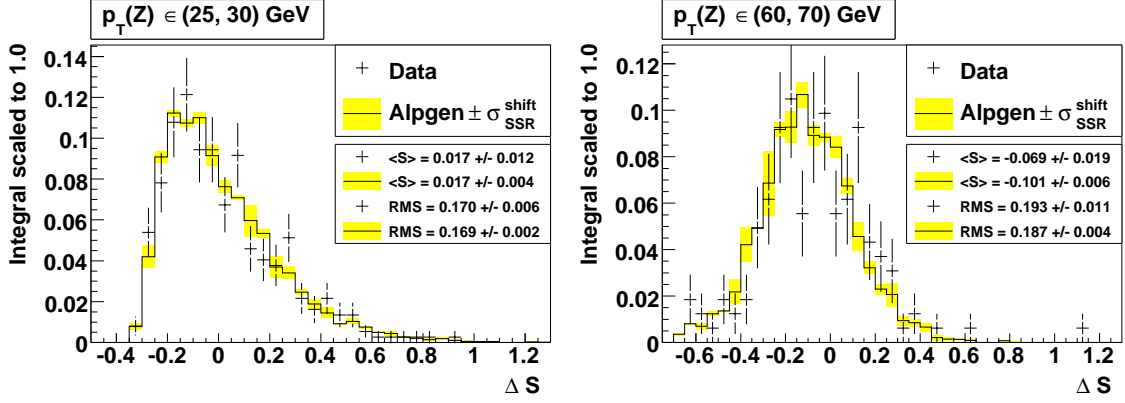


Figure 9.13: Data and simulation for ΔS for the exclusive $Z + 1$ -jet back-to-back sample for p_T^Z in the ranges 25 – 30 GeV (left) and 60 – 70 GeV (right).

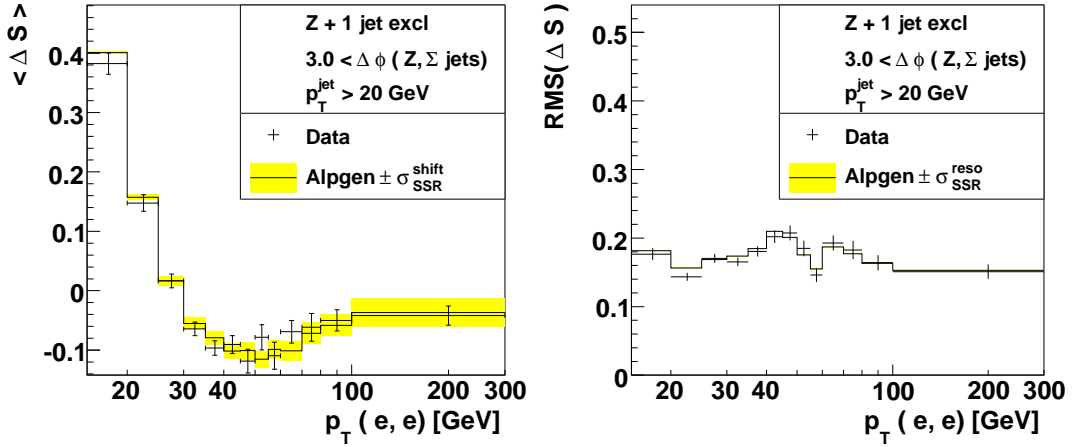


Figure 9.14: Simulation compared with data for the mean (left) and RMS (right) of the ΔS distribution in bins of p_T^Z . Events were required to satisfy $\Delta\phi(Z, \text{leading jet}) > 3.0$ and contain one and only one jet. The yellow bands indicate the systematic uncertainties quoted for the jet energy scale and resolution corrections applied to simulation, whereas the shaded yellow band corresponds to the jet energy scale uncertainties in data.

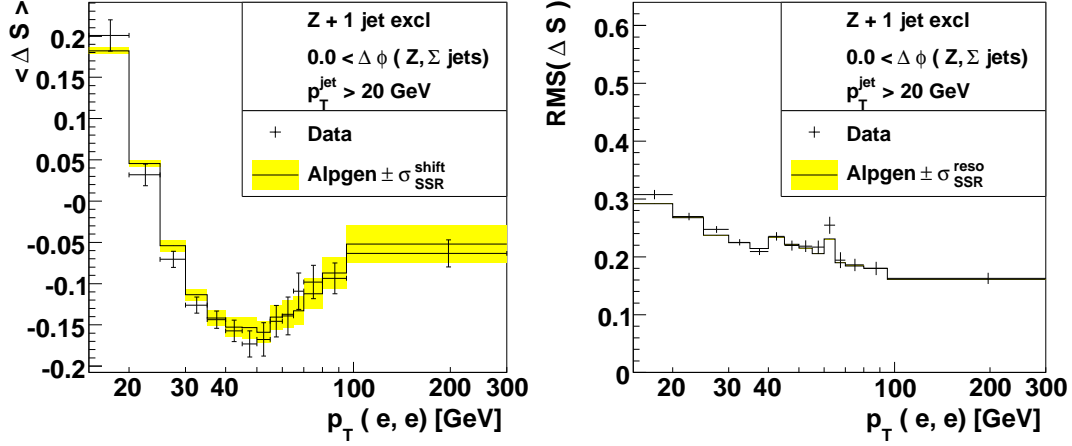


Figure 9.15: Simulation compared with data for the mean (left) and RMS (right) of the ΔS distribution in bins of p_T^Z . Events were required to contain one and only one jet, with no $\Delta\phi(Z, \text{leading jet})$. The yellow bands indicate the systematic uncertainties quoted for the jet energy scale and resolution corrections applied to simulation, whereas the striped shaded yellow band corresponds to the jet energy scale uncertainties in data.

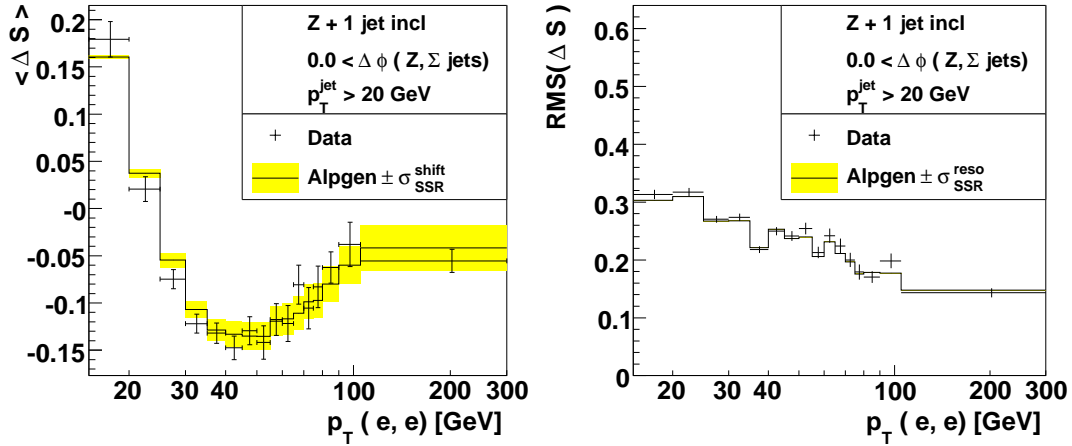


Figure 9.16: Simulation compared with data for the mean (left) and RMS (right) of the ΔS distribution in bins of p_T^Z . Events were required to contain one or more jets, with no $\Delta\phi(Z, \text{leading jet})$. The yellow bands indicate the systematic uncertainties quoted for the jet energy scale and resolution corrections applied to simulation, whereas the shaded yellow band corresponds to the jet energy scale uncertainties in data.

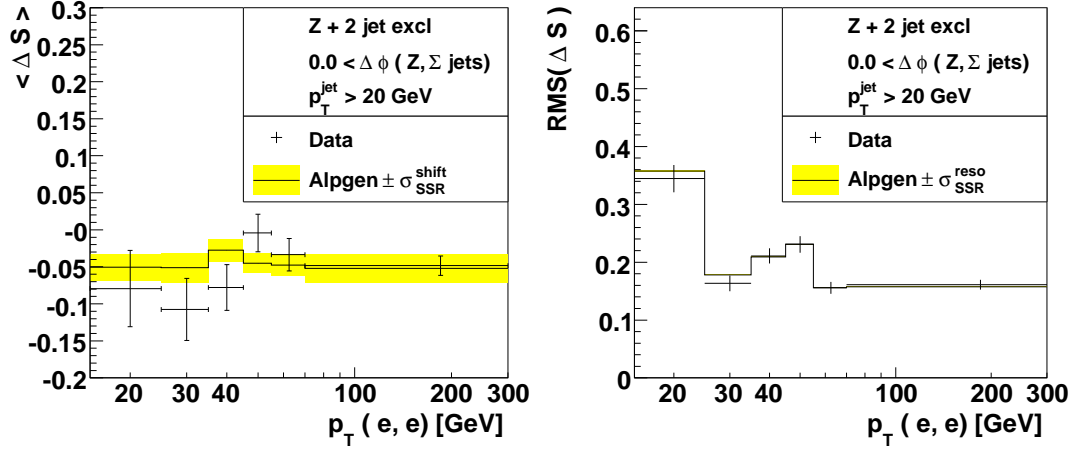


Figure 9.17: Simulation compared with data for the mean (left) and RMS (right) of the ΔS distribution in bins of p_T^Z . Events were required to contain two or more jets, with no $\Delta\phi(Z, \text{leading jet})$. The yellow bands indicate the systematic uncertainties quoted for the jet energy scale and resolution corrections applied to simulation, whereas the shaded yellow band corresponds to the jet energy scale uncertainties in data.

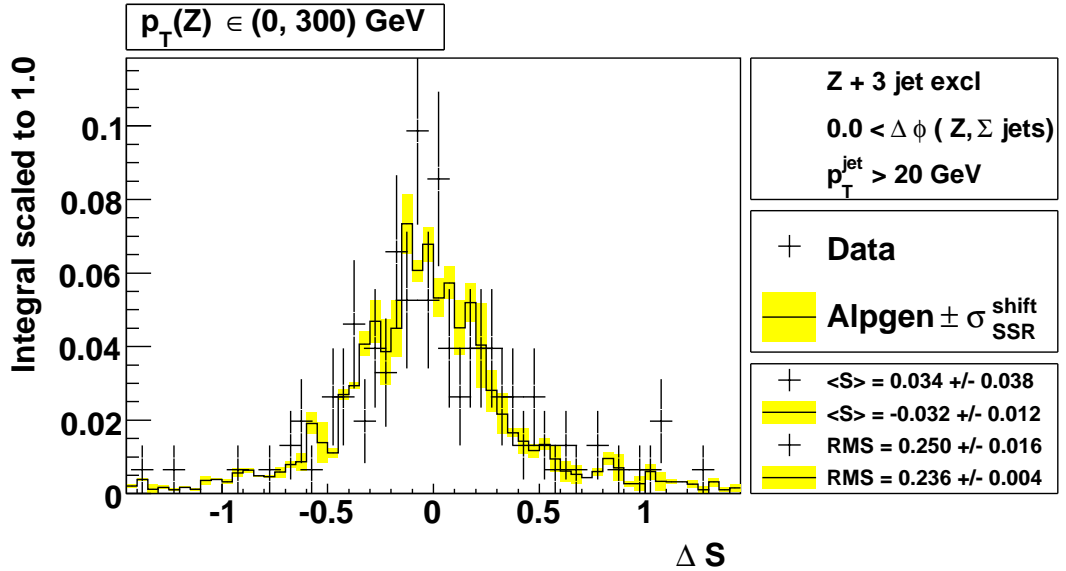


Figure 9.18: ΔS distribution in events with three or more jets for simulation and data. The yellow bands indicate the systematic uncertainties quoted for the jet energy scale corrections applied to simulation.

9.13.1 Jet Dependency of the Trigger Efficiency

The trigger efficiency for jet-inclusive $Z/\gamma^* \rightarrow e^+e^-$ events with $p_T^e > 25$ GeV is known to be $\sim 100\%$ based on a tag-and-probe analysis [89]. The event selection for the efficiency measurement requires $\Delta\phi$ between the two electrons greater than 2.0 to reduce the QCD background. The efficiency of this selection criterion is expected to decrease with increasing p_T^{jet} and jet multiplicity. Since the cross section for the production of $Z/\gamma^*(\rightarrow e^+e^-) + N$ jets is suppressed by approximately α_s^N with respect to the inclusive cross section, the trigger efficiency for events with one or more jets could potentially be significantly less than 100% without being in contradiction with the jet-inclusive measurement.

To test if the trigger efficiency shows a dependency on the number of jets in the event, the data sample passing the normal event selection was selected, without applying any trigger selection. For each event N_{failed} is defined as the number of reconstructed electrons associated with the Z decay which failed to trigger any single-EM trigger. In case that the presence of jet(s) in an event would reduce the single-EM trigger efficiency significantly, N_{failed} would on average be larger in the samples with at least N jets ($N = 1, 2, 3$) than in the inclusive sample ($N = 0$). In Fig. 9.19 the distribution of N_{failed} in data is given for events containing $N = 0, 1, 2, 3$ or more jets, and no indication of a reduction in trigger efficiency as a function of jet multiplicity is seen.

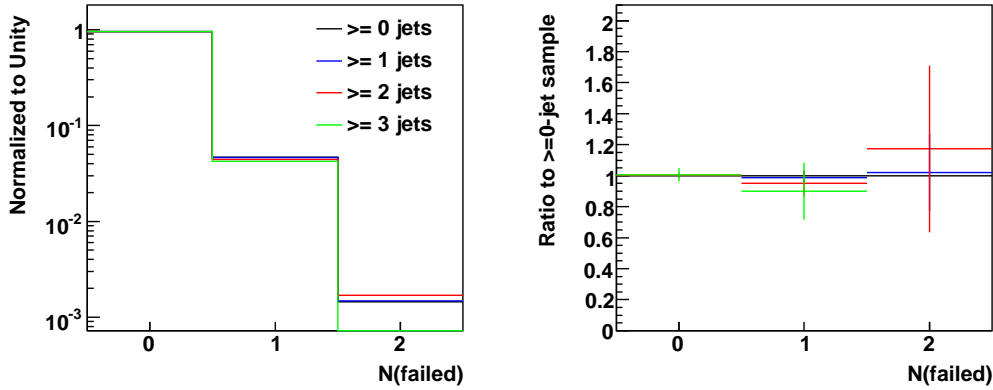


Figure 9.19: The number of Z decay electrons in selected data events which failed to trigger any single-EM triggers as function of the jet multiplicity.

To verify that the electrons which fail to trigger any single-EM trigger do so due to their low p_T , the p_T distributions for all electrons, for electrons which fired at least one single-EM trigger (*triggered electrons*), and for electrons which failed to trigger any single-EM trigger were studied for events with N or more jets, $N = 0, 1, 2, 3$. The p_T ratio of triggered electrons to all electrons for jet-inclusive events ($N = 0$) shows the properties of a turn-on function expected from the minimal E_T criteria of the triggers, as shown in Fig. 9.20. The ratios for $N = 1, 2, 3$ are compatible with that for $N = 0$, validating that jets do not reduce the trigger efficiency³.

³Note that the trigger efficiency is defined for events which pass (would have passed) the electron ID criteria. I.e. if nearby hadrons from jet activity leads to an electron failing both the trigger and ID criteria, this does lead to a reduction in the trigger efficiency.

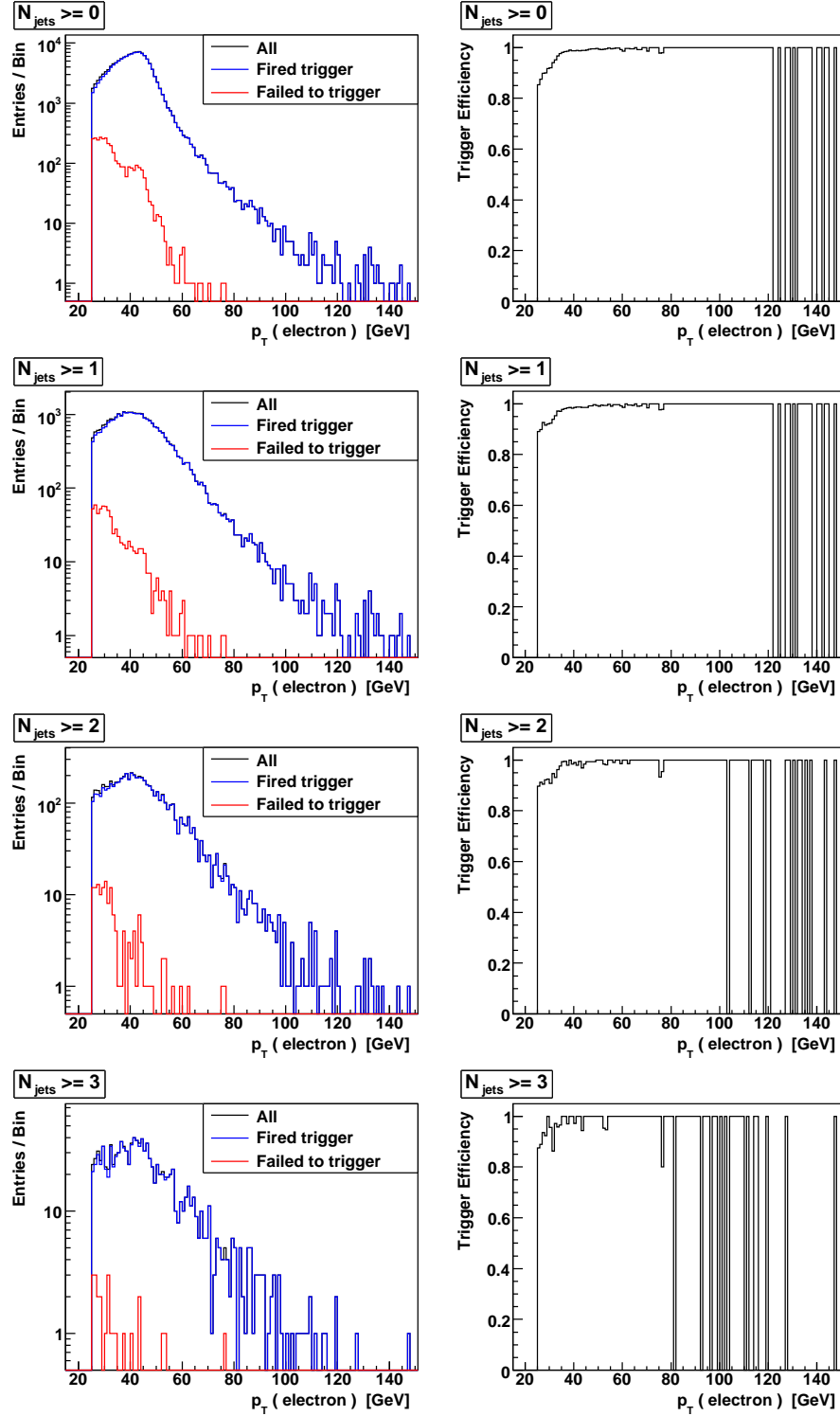


Figure 9.20: Transverse momentum distributions for all electrons, those which triggered a single-EM trigger (*triggered electrons*), and those which did not, for events with, from top to bottom on the left, zero, one, two or three or more jets. The corresponding ratios of the p_T of triggered electrons to the p_T of all electrons are shown on the right. No significant dependency on the jet multiplicity is seen.

9.13.2 Jet Dependency of the Electron ID Efficiency

The electron identification criteria includes several variables which might be influenced by the presence of QCD activity in the event. If hadrons from QCD radiation deposit their energy close to the electron, the electron candidate might fail the ID criteria. If this loss of electrons in data is correctly described in simulation, this effect is taken into account when correcting the measurement to the particle level since $(\mathcal{A} \times \epsilon)_{\text{data}} = (\mathcal{A} \times \epsilon)_{\text{sim}}$. However, since the electron ID efficiencies in data and simulation were measured for jet-inclusive samples, it must be tested to which level of accuracy the correlation between the electron ID efficiency and QCD activity is modeled by the simulation.

Jets pointing in the Direction of Electrons

As a *gedankenexperiment*, let's assume an event containing only a single electron and a single jet, both objects carrying the same p_T , and that such events are available in bins of $\Delta R(e, \text{jet})$ for both data and simulation. For $\Delta R \leq X$ the electron always fails the ID criteria due to the isolation, shower shape and electromagnetic energy fraction criteria. For $X < \Delta R \leq Y$ the electron may or may not fail the ID criteria depending on how broad the jet is. For $\Delta R > Y$ the two objects do not interfere, and the electron ID efficiency is identical to in events with no jet. The average ID efficiency in the sample will depend both on the shape of the $\Delta R(e, \text{jet})$ distribution and on how broad the energy deposit arising from a jet is on average.

It is tested to which extent the $\Delta R(e, \text{jet})$ distribution in the simulated $Z/\gamma^* \rightarrow e^+e^-$ sample agrees with data at detector level. For each event with exactly one jet the electron-jet pair with the smallest value of $\Delta R(e, \text{jet})$ is selected, and the distribution of this minimal ΔR is compared between particle level and detector level in the ALPGEN+PYTHIA sample (see Fig. 9.21 (left column, top)). The two distributions are in good agreements for $\Delta R > 0.8$, but for smaller value a lack of events at detector level is observed, corresponding to events where an electron fails to be identified as such due to the nearby jet. At detector level the integral of the distribution from $\Delta R = 0.0$ to $\Delta R = 0.8$ is 0.3 compared with 0.11 at particle level, indicating that 8% of the events fail the detector-level selection due to interference between electrons and jets. The detector-level distribution in ALPGEN+PYTHIA shows good agreement with data (Fig. 9.21 (left column, center)), both above and below the kink at $\Delta R = 0.8$. No significant differences are seen between the particle-level distribution predicted by ALPGEN+PYTHIA and those predicted by SHERPA and PYTHIA (Fig. 9.21 (left column, bottom)). Similar conclusions can be drawn for events with two jets (see Fig. 9.21 (central column)) and three jets (see Fig. 9.21 (central column)). This shows that the ALPGEN+PYTHIA sample passed through the detector simulation gives an accurate description of the inefficiency of the electrons identification in data due to jets with a $\Delta R(e, \text{jet})$ value of less than 0.8.

Low- p_T Particles pointing in the Direction of Electrons

It could be the case that the inefficiency of the electron identification due to jets satisfying $p_T > 20$ GeV is well described by simulation, but that there are discrepancies between data and simulation for the inefficiency arising from low- p_T QCD radiation which is not clustered in a jet. Due to the energy resolution in the calorimeter, and to the jet reconstruction threshold, this is difficult to study using calorimeter information only. This low- p_T radiation

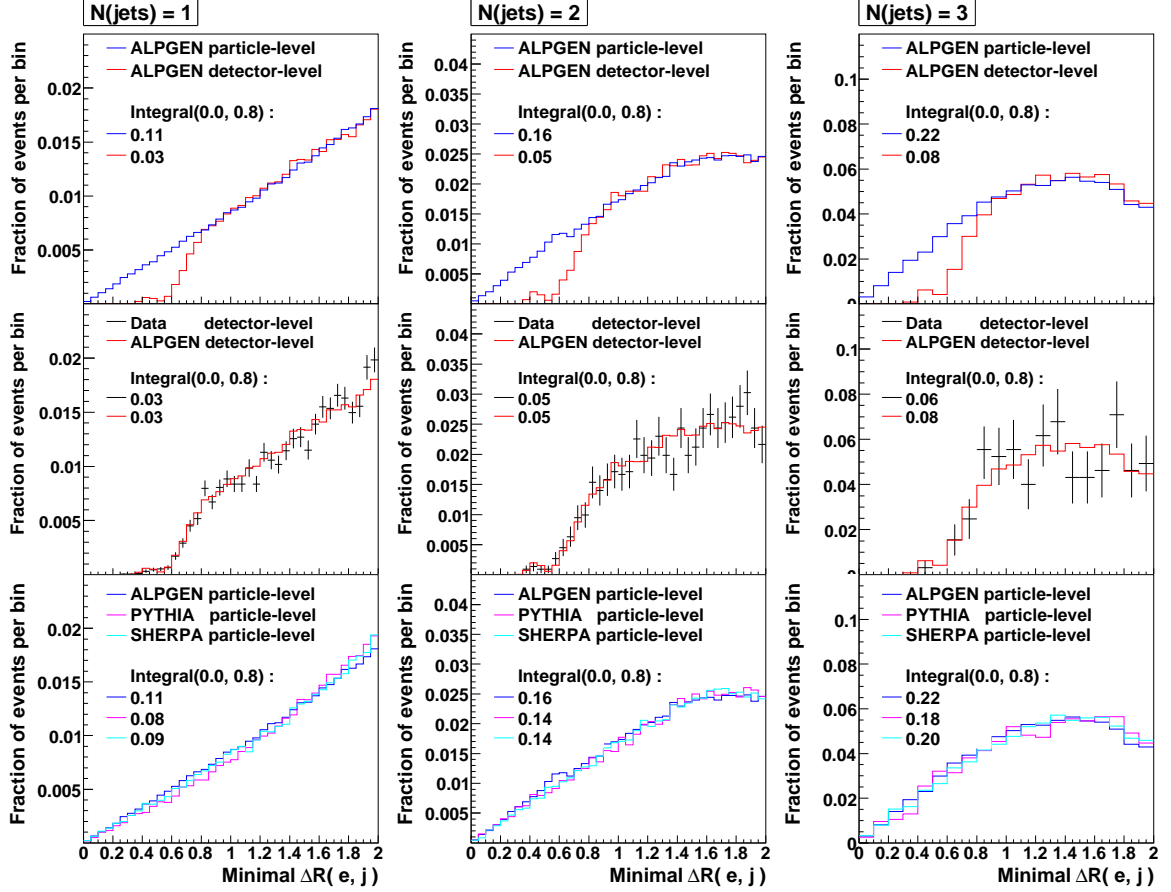


Figure 9.21: Distributions of the minimal $\Delta R(e, \text{jet})$ distance of all electron-jet pairs in event with one jet (left column), two jets (central column) and three jets (right column). The particle-level and detector-level distributions in ALPGEN+PYTHIA (upper row) indicate that electrons tend to fail the ID criteria if $\Delta R(e, \text{jet}) < 0.8$. The integral of each distribution for the interval $0.0 < \Delta R(e, \text{jet}) < 0.8$ is given in the legend. Data and ALPGEN+PYTHIA are in agreement within the statistical uncertainties (central row). No significant differences are seen between the particle-level distributions predicted by ALPGEN+PYTHIA SHERPA and PYTHIA (lower row).

is probed using the multiplicity and momentum distribution of charged tracks in the event. The information on neutral particles is lost, but the momentum resolution is improved as compared with the calorimeter. Tracks are selected through an upper cut on the χ^2 of the fitted track with respect to the hits in the tracker and $|\eta| < 2.5$. Only tracks which point towards the primary interaction vertex are taken into consideration. No p_T cut was applied explicitly, but the reconstruction threshold is 400 MeV. Tracks which are closer to the axis of a reconstructed jet than 0.5 are rejected, as are tracks closer than 0.4 to an electron. The scalar p_T sum and average p_T of tracks passing the selection are studied in bins of jet multiplicity. Assuming that the simulated event sample has a different electron ID efficiency in N -jet events than data due to different amounts of low- p_T particles from QCD activity, the ratio of the distribution of the scalar track p_T sum (or average track p_T) in N -jet events relative to in 0-jet events in simulation must be different than in data.

The comparisons between 1-jet and 0-jet events are given in Fig. 9.22 (left) and Fig. 9.23 (left). Events with one jet have, on average, a significantly higher scalar p_T track sum, and a higher average track p_T , than 0-jet events. While the distributions for 0-jet and 1-jet events in simulation do not agree with data, the ratio is seen to be well described. Studying the ratio between the track distributions in N -jet and 0-jet events cancels any differences between data and simulation arising from effects which are not correlated with the number of jets from the hard scattering, e.g. tracks arising from multiple parton interactions or from additional $p\bar{p}$ collisions during the same bunch crossing. Since the ID efficiency for inclusive events in simulation has already been corrected to data (see Sect. 9.10), it is enough that the simulation describes the track distributions relative to 0-jet events correctly. In addition, due to the normalization of the measurements to the inclusive cross section, only disagreements between data and simulation which correlate with jet activity are of importance. The same distributions, for 2-jet and 3-jet events, are given in Fig. 9.22 (center) and (right) and Fig. 9.23 (center) and (right), respectively. It is seen that simulation and data disagree for the distributions themselves, whereas agreement is seen for the ratios to the distributions for 0-jet events.

9.14 Correcting Kinematic Properties of the Simulated Event Sample

Having applied the corrections above, the performance of the simulated and real detector is equal with respect to all properties judged to be relevant for the measurements of the normalized p_T^{jet} spectra. Comparisons between reconstructed spectra in data and simulation for $p_T(\text{dielectron})$, $p_T(\text{leading jet})$, $p_T(\text{dielectron})$ for 1-jet events and $\Delta\phi(Z, \text{jet})$ are given in Fig. 9.24. The simulated events sample was normalized to the number of jet-inclusive events in data. Significant differences are seen for all four observables. A minimum for $(\mathcal{A} \times \epsilon)_{\text{data}} = (\mathcal{A} \times \epsilon)_{\text{sim}}$ to be valid for the p_T^{jet} spectra which are to be measured is that the p_T^{jet} spectra themselves show agreement. However, in order to claim that the data sample is well understood, reasonable agreement for all four observables would be preferable. Several different reweighting schemes for achieving this were devised and tested.

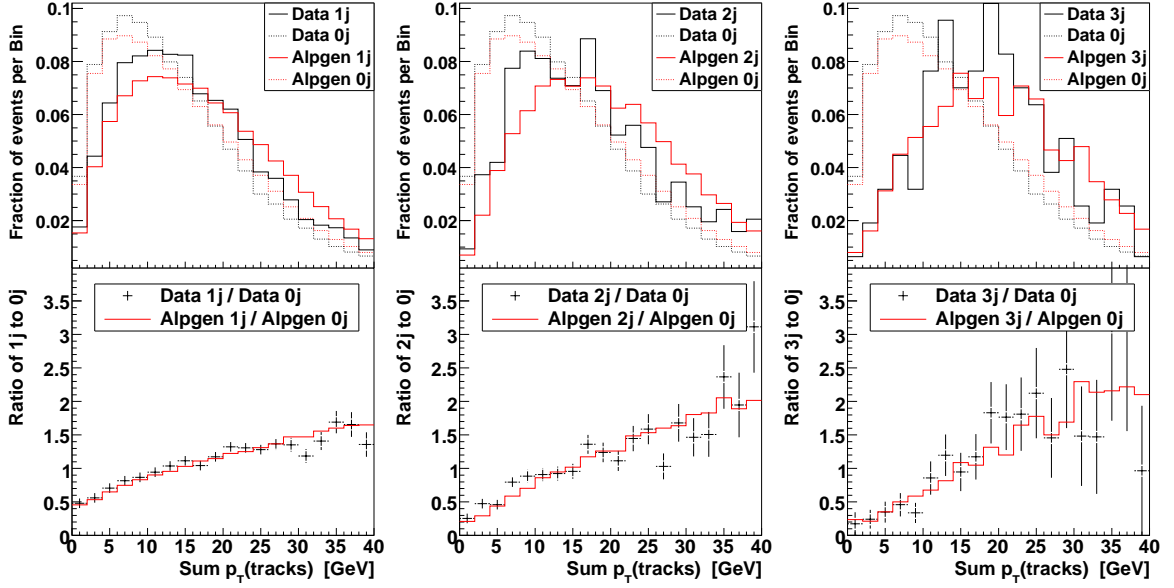


Figure 9.22: The scalar p_T sum of tracks in N -jet and 0-jet events for $N = 1$ (top, left), $N = 2$ (top, center) and $N = 3$ (top, right). Only tracks which are not pointing towards a reconstructed jet or electron were taken into account. The ratios of the N -jet distributions to the 0-jet distributions are shown for $N = 1$ (bottom, left), $N = 2$ (bottom, center) and $N = 3$ (bottom, right).

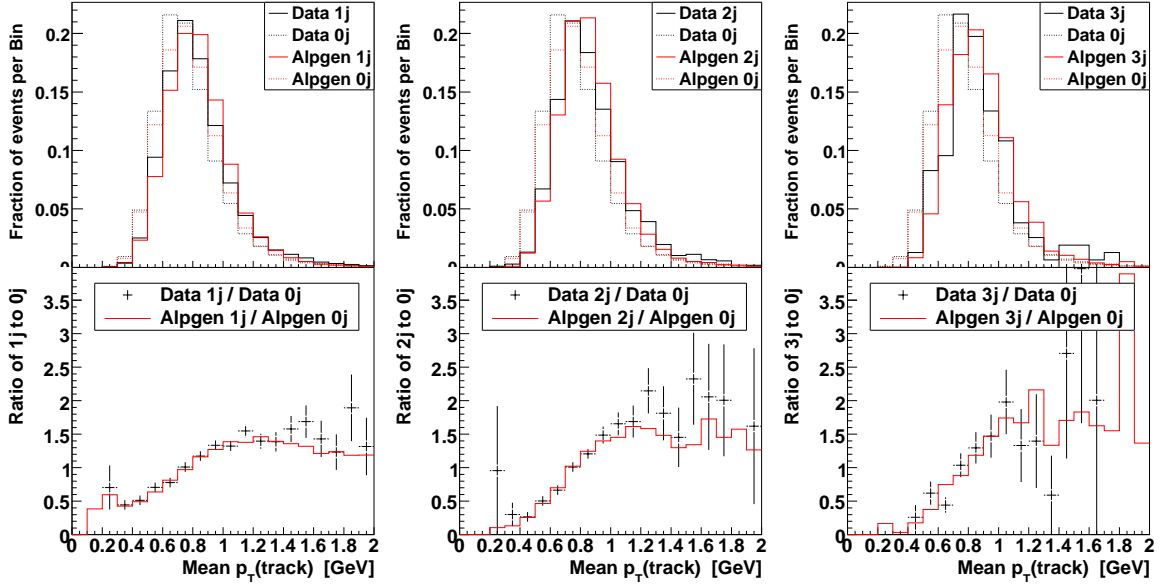


Figure 9.23: The average p_T of tracks in N -jet and 0-jet events for $N = 1$ (top, left), $N = 2$ (top, center) and $N = 3$ (top, right). Only tracks which are not pointing towards a reconstructed jet or electron were taken into account. The ratios of the N -jet distributions to the 0-jet distributions are shown for $N = 1$ (bottom, left), $N = 2$ (bottom, center) and $N = 3$ (bottom, right).

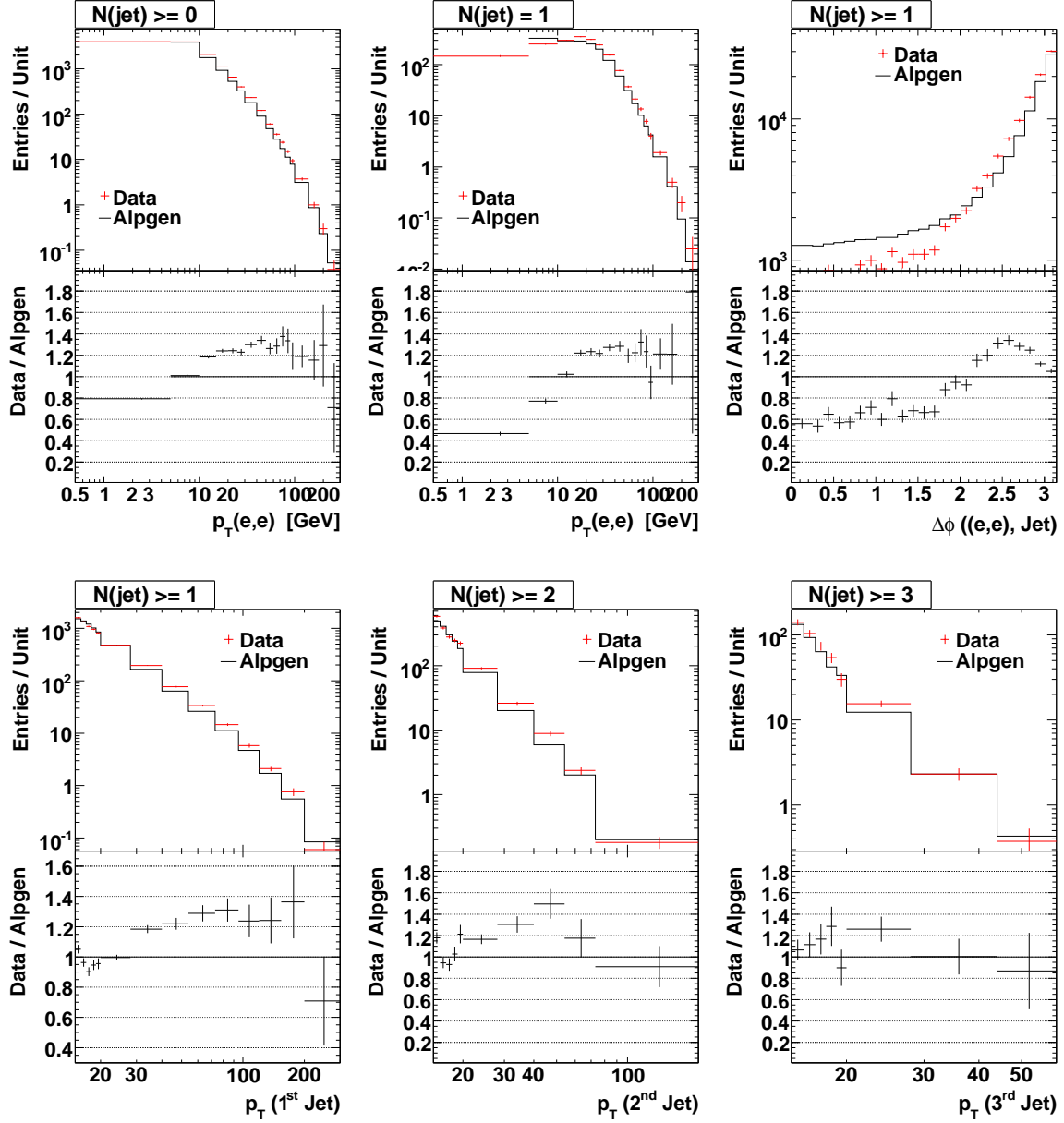


Figure 9.24: Data vs ALPGEN+PYTHIA (labeled *Alpgen*) for the following observables (from top to bottom): p_T^Z for jet-inclusive events, $p_T(\text{leading jet})$ and finally p_T^Z and $\Delta\phi(Z, \text{leading jet})$ for events with at least one jet. The distribution is shown on the left, the ratio data to simulation on the right. All p_T values are given in units of GeV.

1D p_T^Z re-weighting

One option is to reweight the simulated event sample as a function of the particle-level p_T^Z to reproduce the DØ Run II p_T^Z measurement [118]. As seen in Fig. 9.25, this improves the agreement for p_T^{jet} as well as, per construction, the p_T^Z of the jet-inclusive sample. However, the p_T^Z in the 1-jet sample is not improved, indicating that data and ALPGEN+PYTHIA have different correlations between p_T^{jet} and p_T^Z in 1-jet events. Additional discrepancies also remain for the $\Delta\phi(Z, \text{jet})$ spectrum.

2D re-weighting to SHERPA

As an alternative, ALPGEN+PYTHIA was reweighted to SHERPA version 1.0.11 at the particle level. The latter generator is known to give a good description of data (see Chapter 7 and Ref. [119]), and a particle-level reweighting to SHERPA is an attempt to transfer this agreement to the ALPGEN+PYTHIA sample. The two event generators are found to have different p_T^Z vs $p_T(\text{leading jet})$ correlations as seen in Fig. 9.26. For both generators the majority of events are approximately compatible with the LO picture where $p_T^Z = p_T^{\text{jet}}$. The ALPGEN+PYTHIA sample in addition have a non-negligible fraction of 1-jet events where p_T^Z is less than 5 GeV and uncorrelated with p_T^{jet} . This class of events does not appear in the subsample of ALPGEN+PYTHIA events where no multiple parton interactions (MPI) took place (*no-MPI* sample), proving that the jets in question arise from MPI. Events where the MPI model produced one or more parton-parton scatter in addition to the hard interaction were rejected by exploiting that the PYTHIA model used to simulate MPI only generates $2 \rightarrow 2$ scatterings and does not employ a parton shower algorithm for MPI scatterings. As a result, all parton pairs generated by the MPI model are exactly back-to-back in the transverse plane. It was confirmed that no such parton pairs occur in event samples generated with the MPI model turned off, meaning that MPI parton pairs can be tagged uniquely in the event-generator history.

It was also attempted to reweight the ALPGEN+PYTHIA sample two-dimensionally to reproduce the $p_T^{\text{jet}}-p_T^Z$ correlation of the SHERPA sample. This lead to a reasonable level of agreement with data from p_T^Z , p_T^{jet} and p_T^Z for 1-jet events, but the discrepancies in the $\Delta\phi(Z, \text{jet})$ spectrum remained. Motivated by this, the correlation between p_T^Z vs $\Delta\phi(Z, \text{leading jet})$ was studied and found to be sensitive to the modelling of MPI as well, see Fig. 9.27. The ALPGEN+PYTHIA sample is seen to have a large amount of events with low p_T^Z and a jet whose direction of flight is uncorrelated with that of the Z boson, and these events do not appear in the no-MPI sample. For this study particle-level jets with p_T down to 5 GeV are taken into consideration. The ALPGEN+PYTHIA sample was reweighted two-dimensionally at particle level to reproduce the correlation seen in SHERPA. For events with zero jets a one-dimensional p_T^Z reweighting to SHERPA was applied instead. This leads to reasonable agreement with data at detector level for all four test variables, see Fig. 9.28. These studies indicate that ALPGEN+PYTHIA using Tune A predicts more jets from MPI than observed in data, and also that SHERPA describes the amount of MPI in data well.

This is in agreement with the Ref. [120] where it is noted that SHERPA needs significantly fewer MPI interactions per event (1.08 compared with 6.35) than PYTHIA using Tune A in order to reproduce studies of dijet events which measure the number and p_T distribution of charged tracks which have a large ΔR separation to the two jets [121]. It is noted that the MPI model in PYTHIA is made to generate a large average number of MPI scatterings

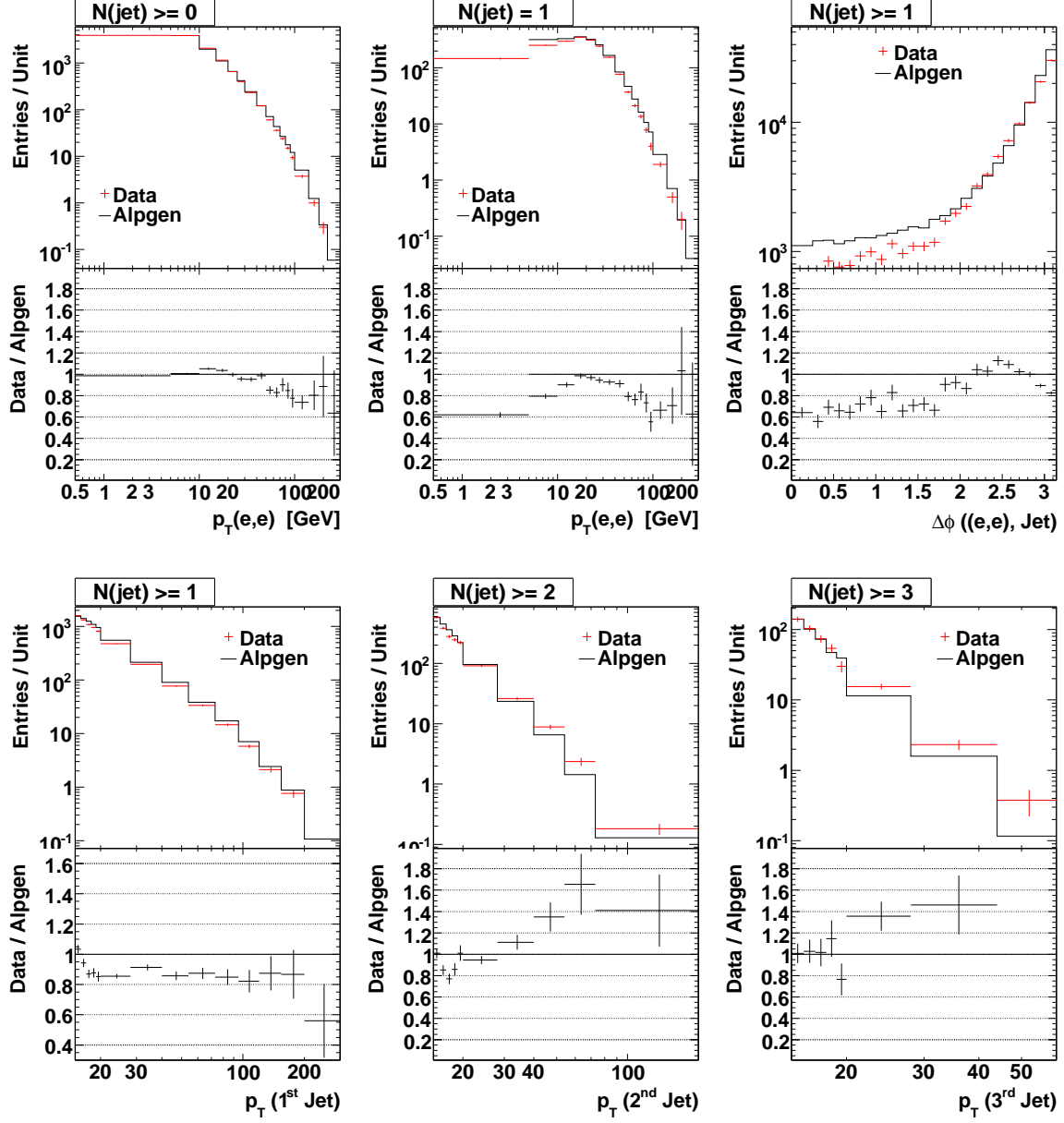


Figure 9.25: Data vs ALPGEN+PYTHIA (labeled *Alpgen*) using p_T^Z -reweighting, for the following observables (from top to bottom): p_T^Z for jet-inclusive events, p_T (leading jet) and finally p_T^Z and $\Delta\phi(Z, \text{leading jet})$ for events with at least one jet. The distribution is shown on the left, the ratio data to simulation on the right. All p_T values are given in units of GeV.

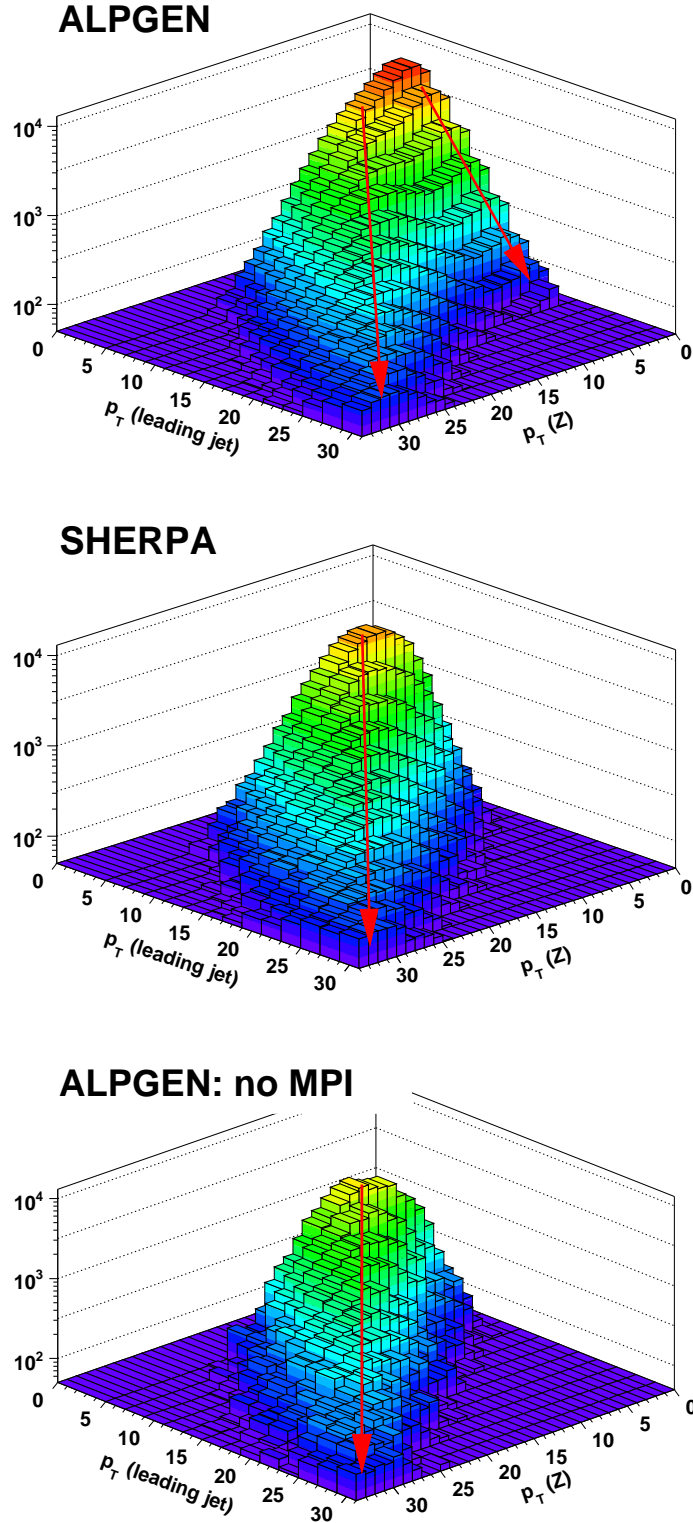


Figure 9.26: The particle-level correlation between $p_T(Z)$ and $p_T(\text{leading jet})$ (both in GeV) for ALPGEN+PYTHIA, SHERPA and the ALPGEN+PYTHIA subsample with no MPI. Local maxima are indicated by red lines.

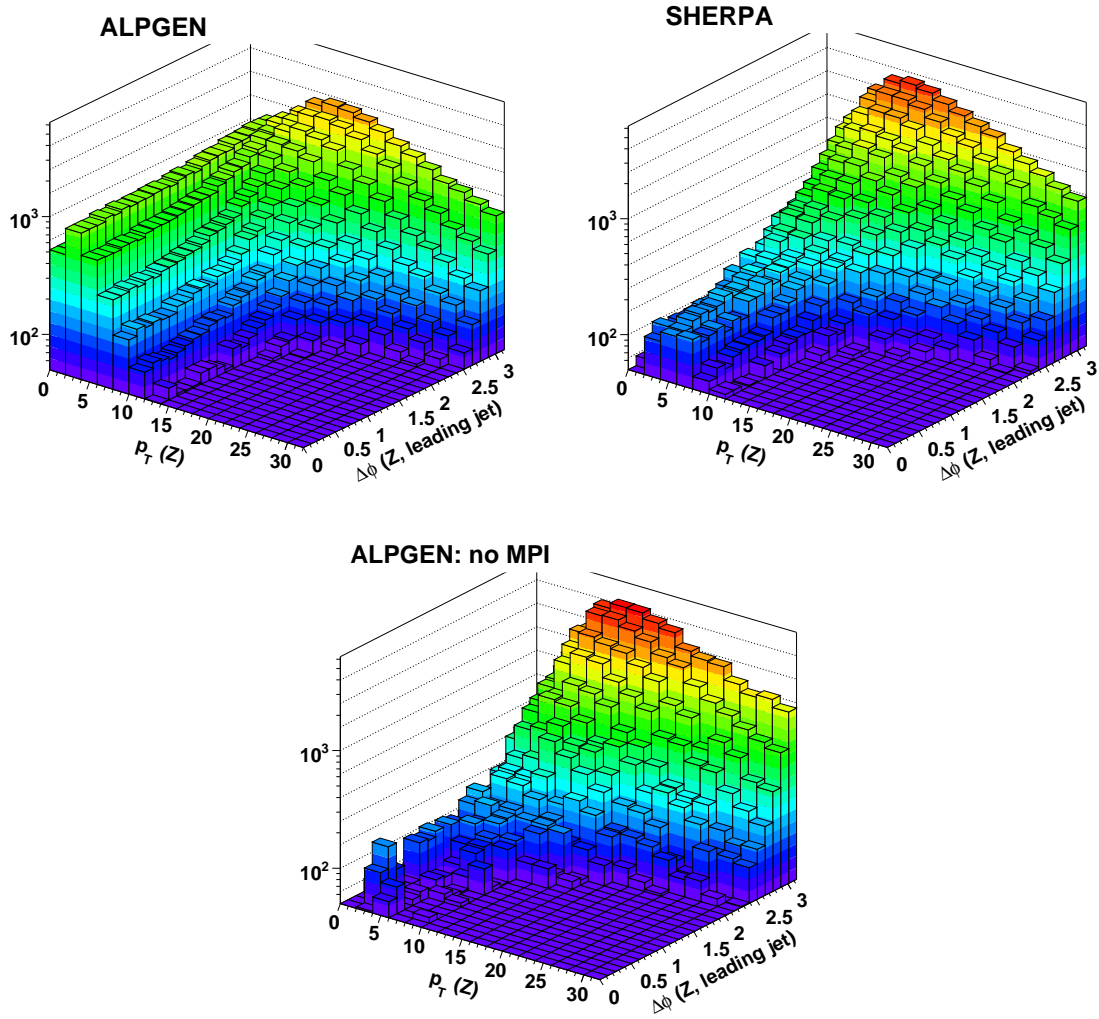


Figure 9.27: The particle-level correlation between $p_T(Z)$ (in GeV) and $\Delta\phi(Z, \text{leading jet})$ for ALPGEN+PYTHIA, SHERPA and the ALPGEN+PYTHIA subsample with no MPI.

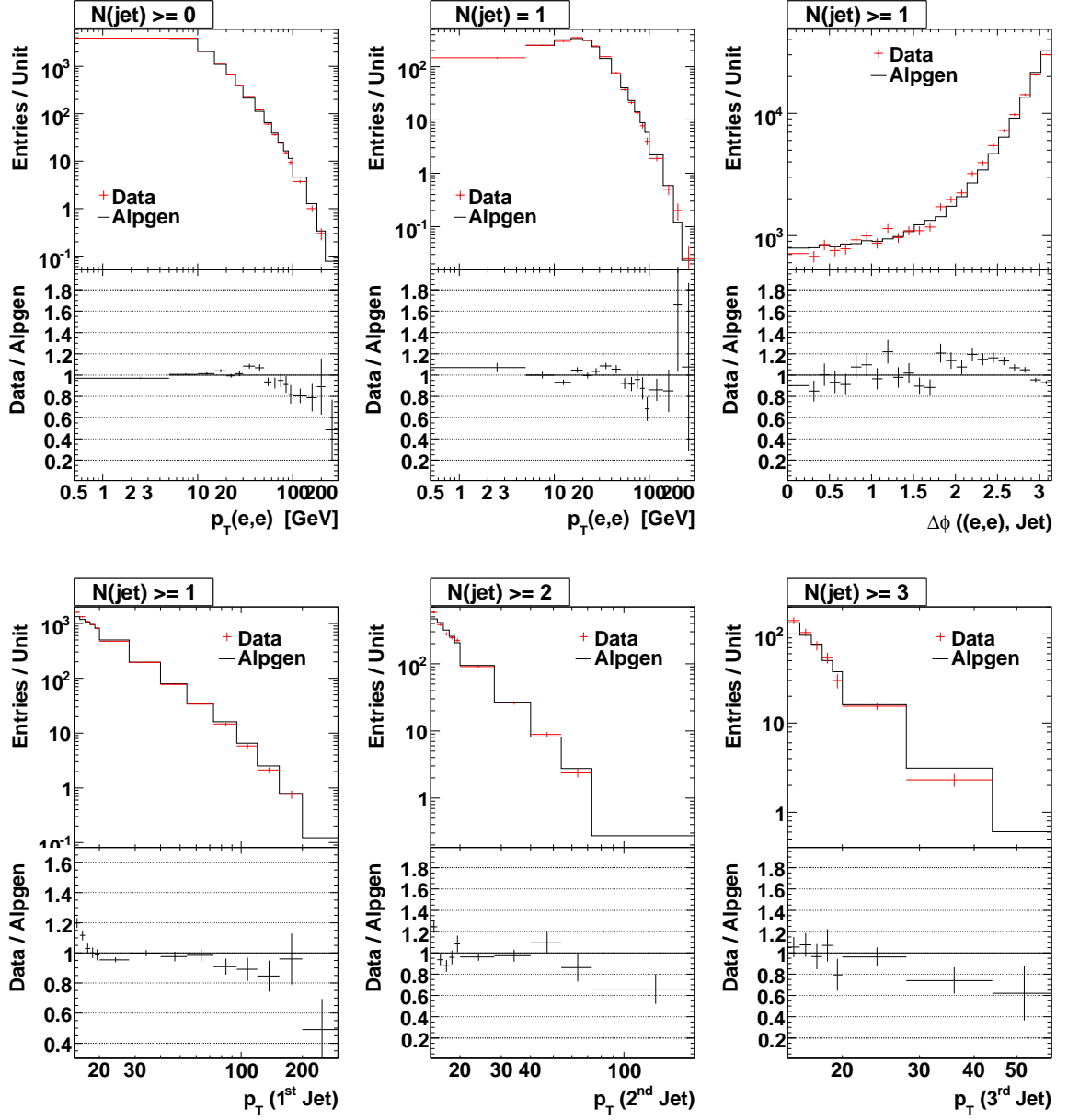


Figure 9.28: ALPGEN+PYTHIA (labeled *Alpgen*) reweighted two-dimensionally in p_T^Z and $\Delta\phi(Z, \text{leading jet})$ to SHERPA compared with data for the following observables (from top to bottom): p_T^Z for jet-inclusive events, $p_T(\text{leading jet})$ and finally p_T^Z and $\Delta\phi(Z, \text{leading jet})$ for events with at least one jet. The distribution is shown on the left, the ratio data to simulation on the right. All p_T values are given in units of GeV.

to compensate for a lack of QCD activity arising from the main parton-parton scattering due to the parton-shower approach used in PYTHIA. In contrast, the matrix-element based description used in SHERPA predicts more tracks arising from the main interaction, leaving less activity to be explained by the MPI model. As for the case of the tuning of intrinsic p_T (see Sect. 3.4.3), this illustrates the danger of absorbing data vs simulation discrepancies arising from compensating for deficiencies in the perturbative, high- p_T calculation using theoretically weakly constrained models which account for low- p_T effects.

The remaining disagreements between the reweighted ALPGEN+PYTHIA sample for the inclusive p_T^Z distribution seen in Fig. 9.28 reflect that SHERPA predicts a somewhat more events with large p_T^Z values than observed in data, as was seen in Sect. 7.3. A second, one-dimensional particle-level reweighting function given by the detector-level ratio of p_T^Z in data to that in the SHERPA-reweighted ALPGEN+PYTHIA sample is performed to correct for this, and the results are given in Fig. 9.29.

Low- p_T bias in simulation

A rapid increase in the ratio data/simulation is seen for $p_T(\text{leading jet})$ below 20 GeV in Fig. 9.28. This effect could arise either due to the event-generator model or to the detector description. In the former case the particle-level $p_T(\text{leading jet})$ spectrum should be reweighted below 20 GeV. Due to the jet energy resolution, this would also impact $(\mathcal{A} \times \epsilon)_{\text{sim}}$ for $p_T^{\text{jet}} > 20$ GeV, thereby influencing the measurements of the p_T^{jet} spectra, where the first measurement bin is 20 – 28 GeV. In case the deviation below 20 GeV arises from a problem with the detector description, only the detector-level spectrum below 20 GeV should be corrected, and the measurements are not influenced. In Fig. 9.30 the reconstructed p_T^{jet} distribution in simulation before jet energy scale corrections (p_T^{uncorr}) is shown for jets with p_T^{jet} after applying jet energy scale and SSR corrections (p_T^{corr}) in the two intervals 15 – 16 GeV and 20 – 22 GeV. The cut-off at 6 GeV reflects the p_T threshold applied in the jet reconstruction algorithm, and the sharp turn-off in the spectrum of p_T^{uncorr} for jets satisfying $15 < p_T^{\text{corr}} < 16$ GeV indicates that jets which failed the 6 GeV threshold would have had a p_T^{corr} value in this range. In data the smearing which is emulated in the SSR method by a Gaussian smearing takes place in the detector, i.e. before the reconstruction threshold is applied, whereas in simulation it happens afterwards, and this difference leads to a bias in the reconstructed p_T^{jet} spectrum in simulation. To evaluate the size of the bias, the p_T^{uncorr} distribution is fitted with a Gaussian, and the size of the missing contribution from jets which failed the threshold cut is estimated as the ratio of the integral of the fit for the interval $[0, 6]$ GeV to the integral for the interval $[0, \infty)$ GeV, yielding 24.2%. For $20 < p_T^{\text{corr}} < 22$ GeV the loss is less than 2%. It was found that simulation agrees well with data down to 15 GeV if the reconstructed $p_T(\text{leading jet})$ spectrum in simulation is corrected for the estimated contribution from jets which failed the reconstruction threshold. Based on this, and that the measurements are performed for $p_T^{\text{jet}} > 20$, no corrections are applied to the simulated event sample to remove the discrepancies seen in Fig. 9.29 below 20 GeV.

1D p_T^{jet} re-weighting to data

So far the reweighting procedure has dealt with discrepancies only for the leading jet. For the measurement, agreement is also needed for the p_T spectra of the 2nd and 3rd jets. Since different slopes translates into different $(\mathcal{A} \times \epsilon)$ separate one-dimensional particle-level reweighting

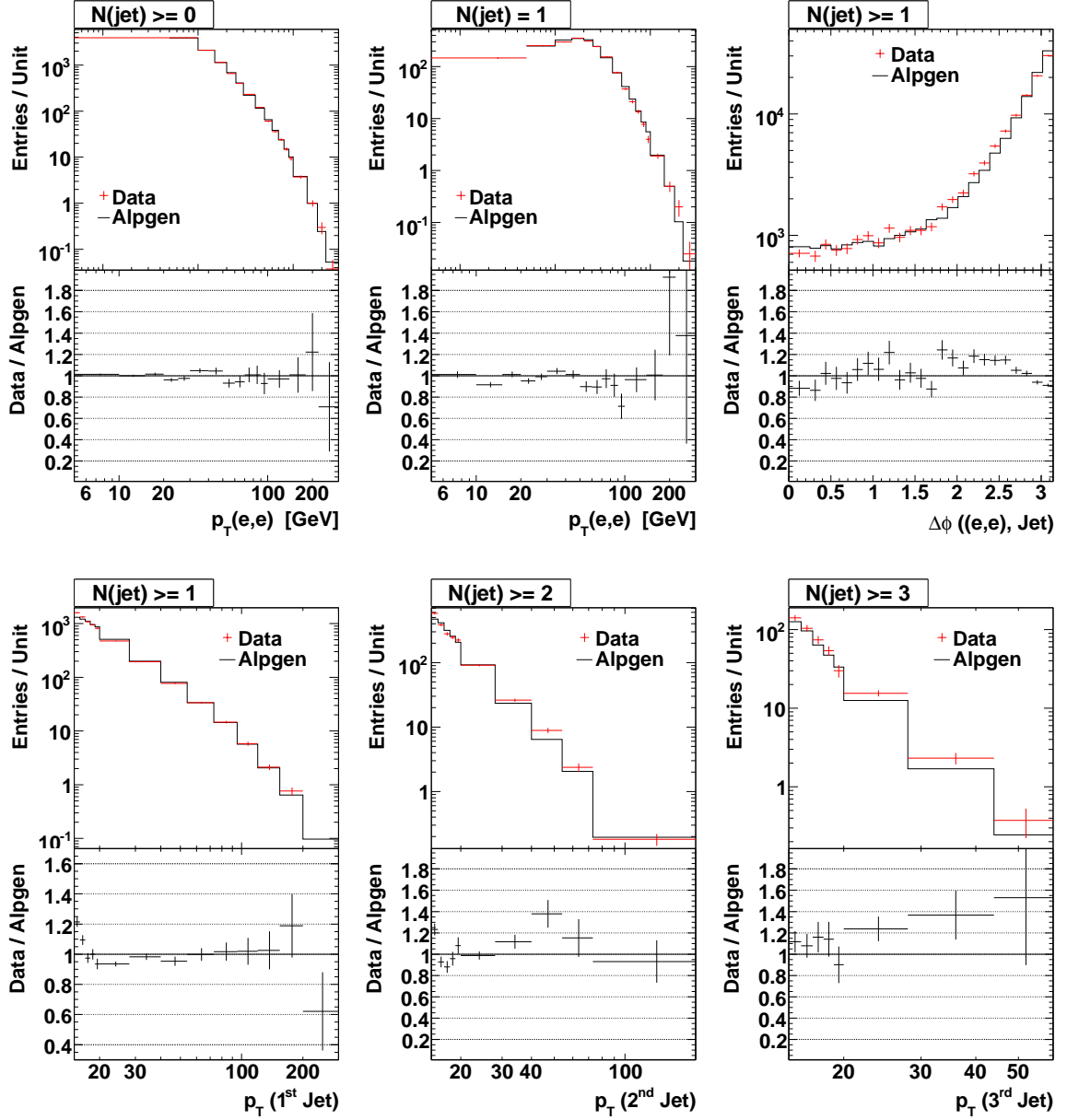


Figure 9.29: Comparisons between data and ALPGEN+PYTHIA (labeled *Alpgen*) reweighted two-dimensionally in p_T^Z and $\Delta\phi(Z, \text{leading jet})$ to SHERPA, and then one-dimensionally in p_T^Z to data. The following observables are shown (from top to bottom): p_T^Z for jet-inclusive events, $p_T(\text{leading jet})$ and finally p_T^Z and $\Delta\phi(Z, \text{leading jet})$ for events with at least one jet. The distribution is shown on the left, the ratio data to simulation on the right. All p_T values are given in units of GeV.

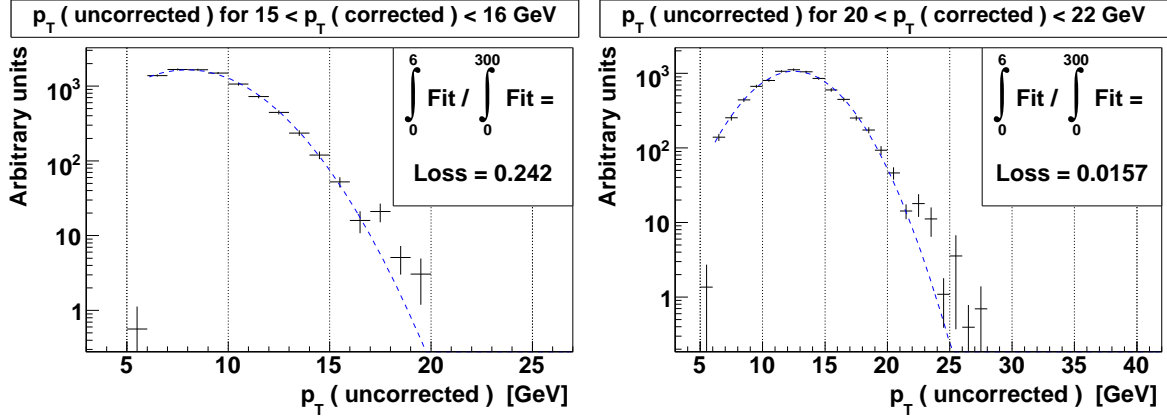


Figure 9.30: The distribution of p_T^{jet} before jet energy scale corrections in bins of the jet energy scale and SSR corrected p_T^{jet} . The contribution which would have come from jets which failed the 6 GeV reconstruction threshold is estimated from the Gaussian fit. For p_T^{corr} in the range 15 – 16 GeV (20 – 22 GeV) the bias is estimated to be $\sim 24\%$ (less than 2%).

functions of the $p_T(N^{\text{th}} \text{ jet})$ to data are applied when measuring $p_T(N^{\text{th}} \text{ jet})$. The ratio of data to simulation for $p_T(N^{\text{th}} \text{ jet})$ before reweighting together with the fitted reweighting function and the ratio after reweighting are given in Fig. 9.31. The difference in $(\mathcal{A} \times \epsilon)$ between applying and not applying the p_T^{jet} reweighting is taken as a systematic uncertainty of the measurement.

1D η^{jet} re-weighting to data

ALPGEN+PYTHIA predicts a narrower η^{jet} distribution than seen in data. To correct for this the Alpgen sample was reweighted as a function of η at particle level using the reweighting functions from Ref [122]. The impact of this reweighting on $\eta(\text{leading jet})$ is seen in Fig. 9.32. The difference in $(\mathcal{A} \times \epsilon)$ between applying and not applying the η^{jet} reweighting is taken as a systematic uncertainty of the measurement.

9.15 Jet Energy Resolution Changes the p_T^{jet} Hierarchy

The p_T^{jet} reweighting to data is defined separately for each of the $p_T(N^{\text{th}} \text{ jet})$ spectra, and agreement between data and simulation is therefore not guaranteed for the p_T spectra of the $(N - 1)^{\text{th}}$ or $(N + 1)^{\text{th}}$ jet. Due to the jet-energy resolution, the N^{th} jet at particle level might be reconstructed as the $(N - 1)^{\text{th}}$ or $(N + 1)^{\text{th}}$ jet at detector level. How often this occurs depends on the relative cross sections $(p_T(N^{\text{th}} \text{ jet})/p_T([N + 1]^{\text{th}} \text{ jet}))$ and $(p_T(N^{\text{th}} \text{ jet})/p_T([N - 1]^{\text{th}} \text{ jet}))$. In case these ratios are significantly different in data and simulation the impact of the jet energy resolution causing the jets to have a different p_T ordering at detector level than at particle level is not described correctly in simulation, with the result that $(\mathcal{A} \times \epsilon)_{\text{data}} \neq (\mathcal{A} \times \epsilon)_{\text{sim}}$. As a cross-check, the ratios are given in Fig. 9.33 for $N = 1, 2, 3$. Reasonable agreement between simulation and data is seen in all cases. Given that the jet

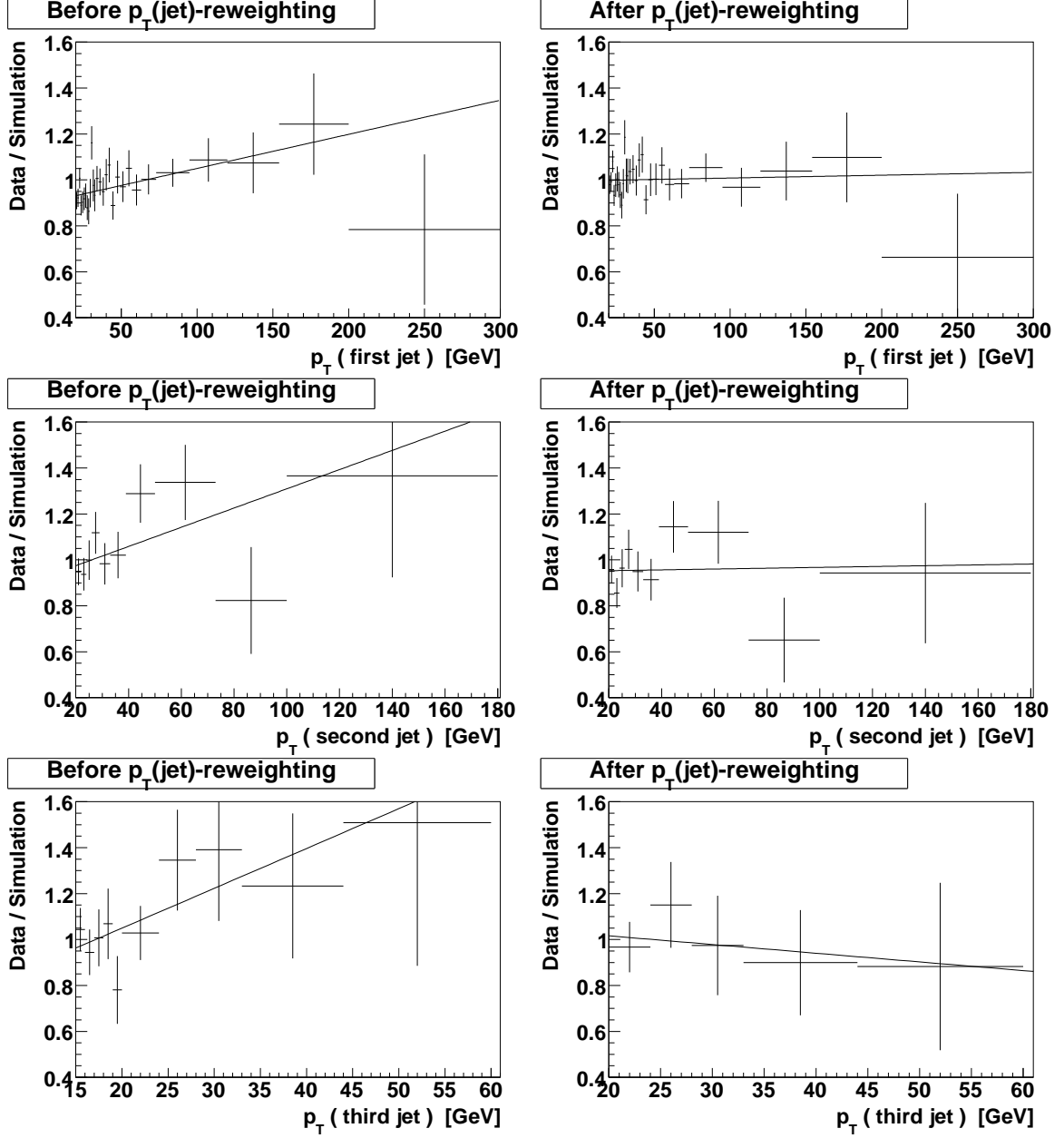


Figure 9.31: The ratios of data to ALPGEN+PYTHIA for $p_T(N^{\text{th}} \text{ jet})$, after reweighting two-dimensionally to SHERPA in p_T^Z and $\Delta\phi(Z, \text{leading jet})$, and one-dimensionally to data in p_T^Z , fitted with a reweighting function (left). Ratios after this additional one-dimensional p_T^{jet} reweighting to data are shown on the right.

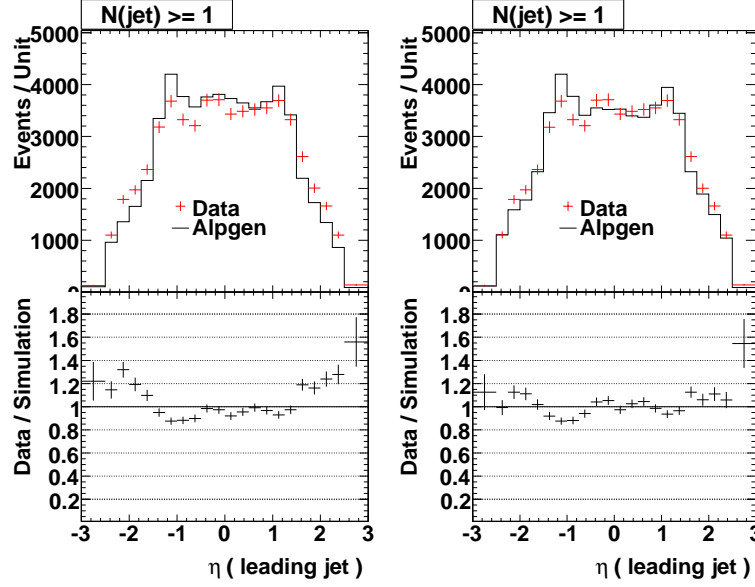


Figure 9.32: Simulation compared with data for $\eta(\text{leading jet})$ before (left) and after (right) applying the η^{jet} -reweighting from Ref. [122].

energy resolution is equal in data and simulation after corrections, the p_T -ordering of the jets will differ between particle-level and detector-level in the same way in simulation and data.

9.16 Total correction factors

Having applied three different categories of corrections to the simulation, accounting for a non-perfect simulation of the collider setup, the detector and the physics of $p\bar{p}$ collisions, $(\mathcal{A} \times \epsilon)_{\text{sim}}$ is equal $(\mathcal{A} \times \epsilon)_{\text{data}}$ within the estimated uncertainties. The systematic uncertainties of the measurements arise from the uncertainties of the background estimates, which were found to be negligible (see Sect. 9.3), and from the uncertainties of the corrections applied to the simulated event sample. Each correction was varied separately within its uncertainties, and the resulting variations in the measured p_T^{jet} spectra were added in quadrature to give the total systematic uncertainty of the measurements. The central values for $(\epsilon \times \mathcal{A})_{\text{sim}}$ for $p_T(\text{three leading jets})$ are shown in Fig. 9.34, together with the predictions corresponding to $\pm 1\sigma$ shift of the main systematic uncertainty of the measurements, namely the uncertainty of the corrections of the jet energy scale in simulation to correspond to that in data.

9.17 Extrapolation for the $\Delta R(e, \text{jet}) > 0.4$ cut

The event selection requires jets to be separated from the two selected electrons by $\Delta R(e, \text{jet}) > 0.4$ both at detector and particle level. The efficiency of this requirement at particle level as determined in event samples generated using ALPGEN+PYTHIA, PYTHIA and SHERPA is given in Fig. 9.35. The measured particle-level spectrum in data is scaled with the inverse of the average of the three efficiency predictions, with the maximal difference between the average and each of the three predictions being assigned as a systematic uncertainty. Since

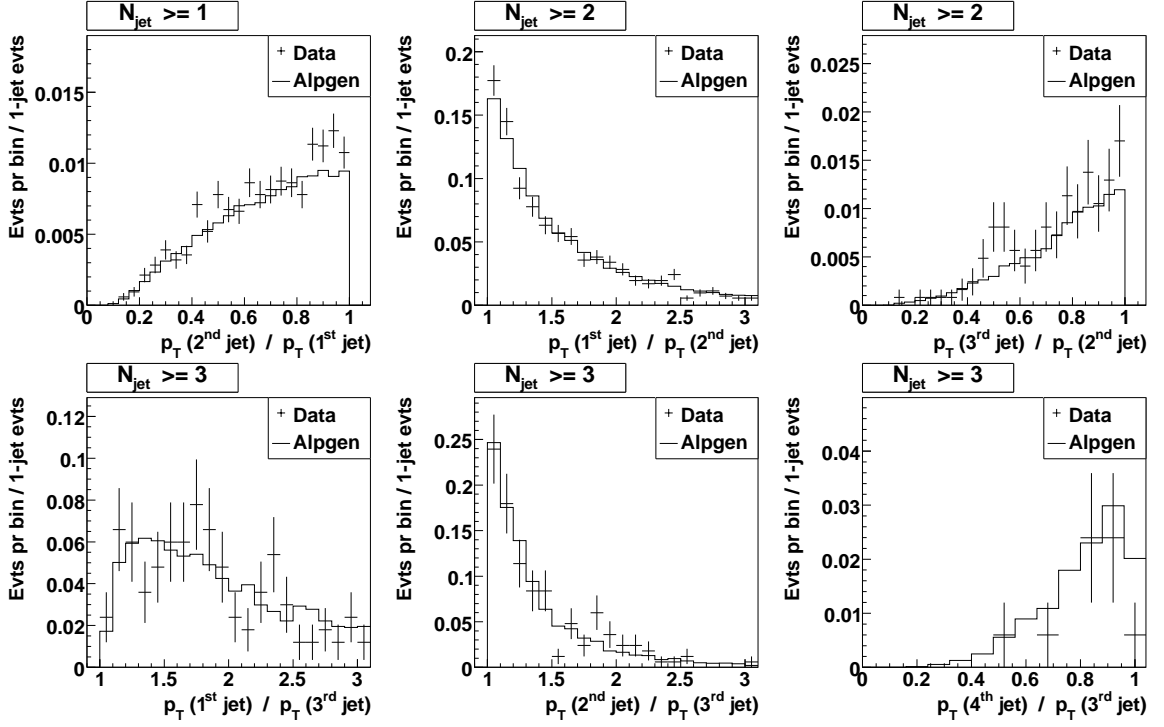


Figure 9.33: The ratios of $p_T(M^{\text{th}} \text{ Jet})$ to $p_T(N^{\text{th}} \text{ Jet})$ for events with at least N jets for $(M, N) = (2, 1)$ (upper, left), $(1, 2)$ (upper, center), $(3, 2)$ (upper, right), $(1, 3)$ (lower, left), $(2, 3)$ (lower, center) and $(4, 3)$ (lower, right). The distributions are normalized to the number of events with at least N jets.

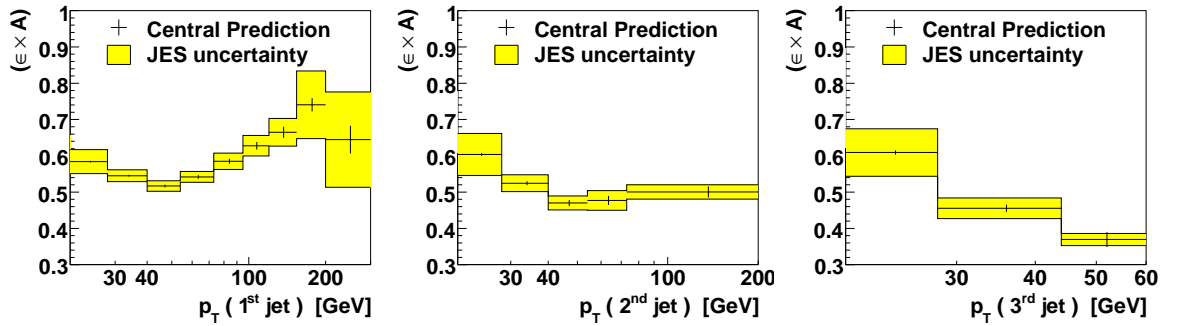


Figure 9.34: $(\mathcal{A} \times \epsilon)_{\text{sim}}$ for the p_T spectra of the first (upper row), second (middle row) and third jet (lower row). The central predictions are shown as points, the prediction corresponding to the relative jet energy scale shifted 1σ up and down are represented by the yellow band.

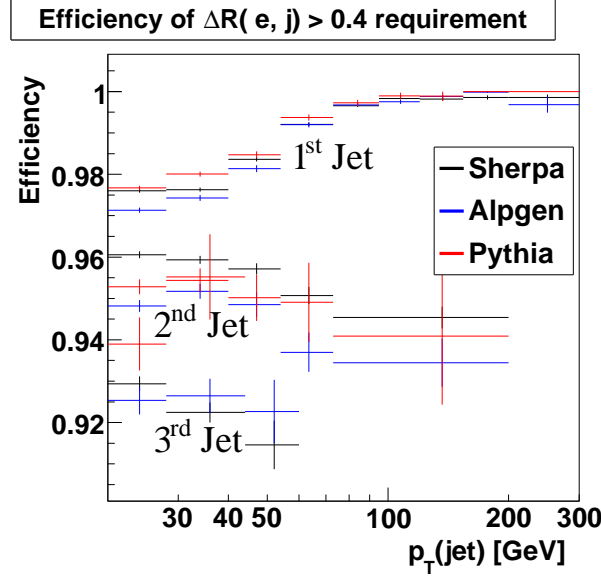


Figure 9.35: The efficiency of the particle-level $\Delta R(e, \text{jet}) > 0.4$ requirement as predicted by ALPGEN+PYTHIA, PYTHIA and SHERPA for the three leading p_T^{jet} spectra.

this uncertainty is below 1% in each bin it is considered to be negligible compared with the main systematic uncertainties of the measurement and is not taken further into account.

9.18 Extrapolation for the lepton cuts

Like the measurement of the inclusive $Z/\gamma^* \rightarrow e^+e^-$ cross section in Sect. 9.11, the p_T^{jet} measurements are performed for two different particle-level selections: once with electron selection criteria corresponding to those applied at detector level (selection *a*), i.e. $p_{T1} > 25$ GeV, $|y^e| < 1.1$ or $1.5 < |y^e| < 2.5$ and $65 < M_{ee} < 115$ GeV and once where only $65 < M_{ee} < 115$ GeV is required (selection *b*). For each particle-level electron selection, the p_T^{jet} spectra are normalized to the inclusive $Z/\gamma^*(\rightarrow e^+e^-) + X$ cross section measured using the same particle-level selection. According to the results presented in Sect. 9.11, the ratio of the inclusive cross section for selection *a* to for selection *b* is 0.443. In Fig. 9.36 this ratio is given as a function of p_T^{jet} for the three leading jets, as predicted by SHERPA, ALPGEN+PYTHIA and PYTHIA. For ALPGEN+PYTHIA the extrapolation was derived both with and without reweighting two-dimensionally to SHERPA in p_T^Z and $\Delta\phi(Z, \text{leading jet})$ as well as one-dimensionally to data in p_T^Z . For PYTHIA the correction was derived with and without a one-dimensional reweighting to data for p_T^Z [123]. The SHERPA prediction is taken as central value and the maximal difference to the four other predictions is quoted as a systematic uncertainty for the measurements for selection *b*. The PDF uncertainty was evaluated using the Hessian method [124] and was found to be negligible. The motivation for extrapolation the measurement from selection *a* to *b* is that it simplifies direct comparisons with other measurements which have a different detector-level lepton acceptance. Originally a combination of the measurements presented in this Chapter with a similar measurement performed in the

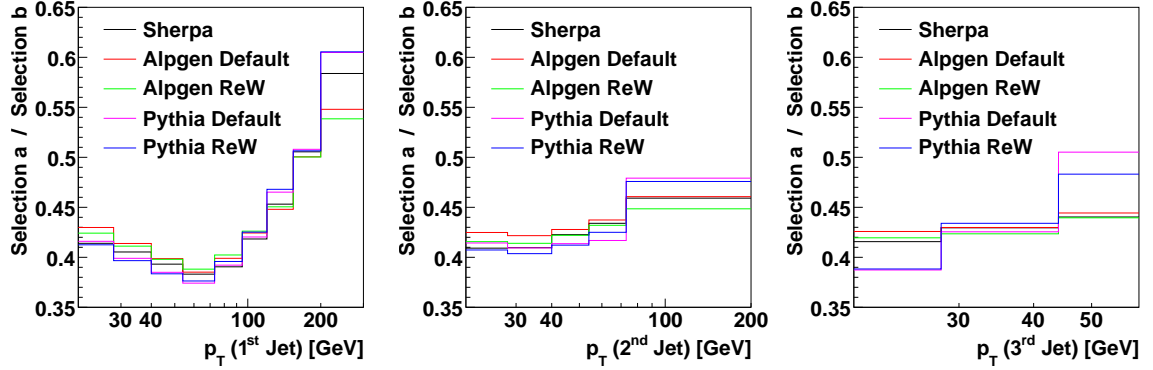


Figure 9.36: Ratio of $p_T(N^{th} \text{ jet})$ for selection a to for selection b as predicted by SHERPA, ALPGEN+PYTHIA and PYTHIA for $N = 1$ (above), $N = 2$ (center) and $N = 3$ (below).

$Z \rightarrow \mu\mu$ channel was foreseen⁴, and since the detector acceptance for electrons and muons differ an extrapolation to a common particle-level selection would be needed to combine the results. This common phase space was foreseen to be selection b .

9.19 Uncertainties

To estimate the uncertainty of the measurement, each correction applied to the simulated event sample was varied separately by $\pm 1\sigma$ and the resulting variations in the measured $\frac{1}{\sigma_{Z/\gamma^*}} \times \frac{d\sigma}{dp_T}$ spectra were added in quadrature to give the total systematic uncertainty of the measurements. The resulting systematic uncertainties for the p_T spectra of the three leading jets are given graphically in Figs. 9.37–9.39 and numerically in Tabs. 9.6–9.7.

The largest source of uncertainty is the correction of the jet energy scale in the simulated event sample to correspond to the jet energy scale of the data sample, contributing (50–80)% of the total systematic uncertainty of the measurements. The uncertainties of the kinematic reweighting of the ALPGEN+PYTHIA sample is taken to be the variation in the measurements arising from applying and not applying the p_T^{jet} and η^{jet} reweighting functions. Additional sources of uncertainties are the corrections of the jet energy resolution and of the jet identification efficiency. The measurements are quoted both for particle-level selection a and b , see Sect. 9.18. For the latter selection the event-generator based extrapolation from selection a to selection b introduces an additional uncertainty, see Sect. 9.18. Presenting the measurements as ratios to the inclusive $Z/\gamma^*(\rightarrow e^+e^-) + X$ cross section cancels the dependence on the uncertainty in the integrated luminosity of the data set.

9.20 Final Results

Having preformed the above-mentioned corrections to the simulated events sample, the particle-level observables in data can be expressed as $O_{\text{part}}^{\text{data}} = (\mathcal{A} \times \epsilon)_{\text{sim}} \times O_{\text{part}}^{\text{data}}$. The

⁴This combination has not been performed since the muon analysis [125] has not been finalized.

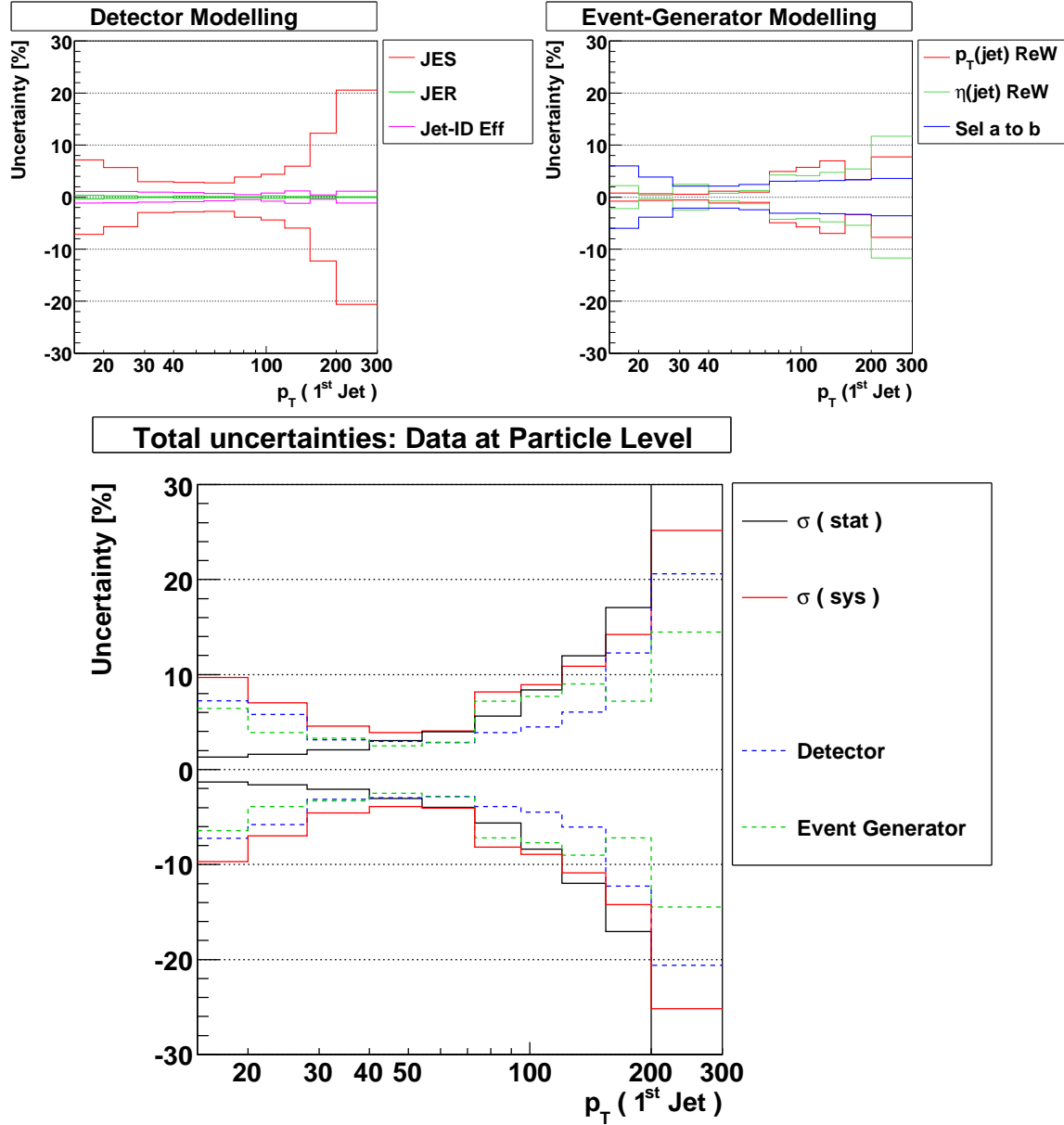


Figure 9.37: The systematic uncertainty of the measurement of $p_T(\text{leading jet})$ is split in two groups of contributions arising from the detector simulation (upper left) and from the event-generator simulation (upper right). In the lower figure the total statistical and systematic uncertainties are given, together with the sum in quadrature of the systematic uncertainties arising from detector and event-generator related corrections. All p_T^{jet} values are in units of GeV.

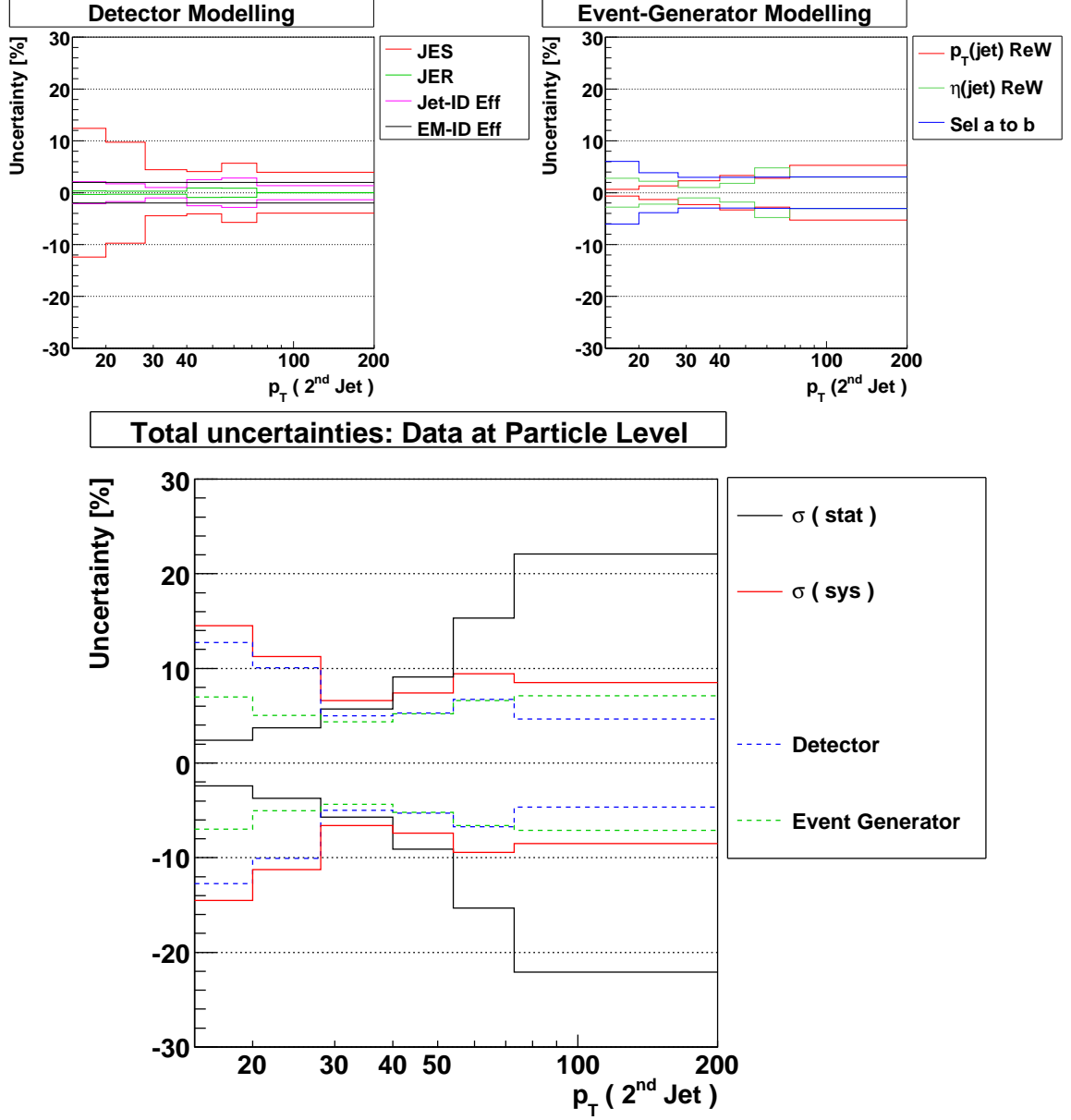


Figure 9.38: The systematic uncertainty of the measurement of $p_T(\text{second jet})$ is split in two groups of contributions arising from the detector simulation (upper left) and from the event-generator simulation (upper right). In the lower figure the total statistical and systematic uncertainties are given, together with the sum in quadrature of the systematic uncertainties arising from detector and event-generator related corrections. All p_T^{jet} values are in units of GeV.

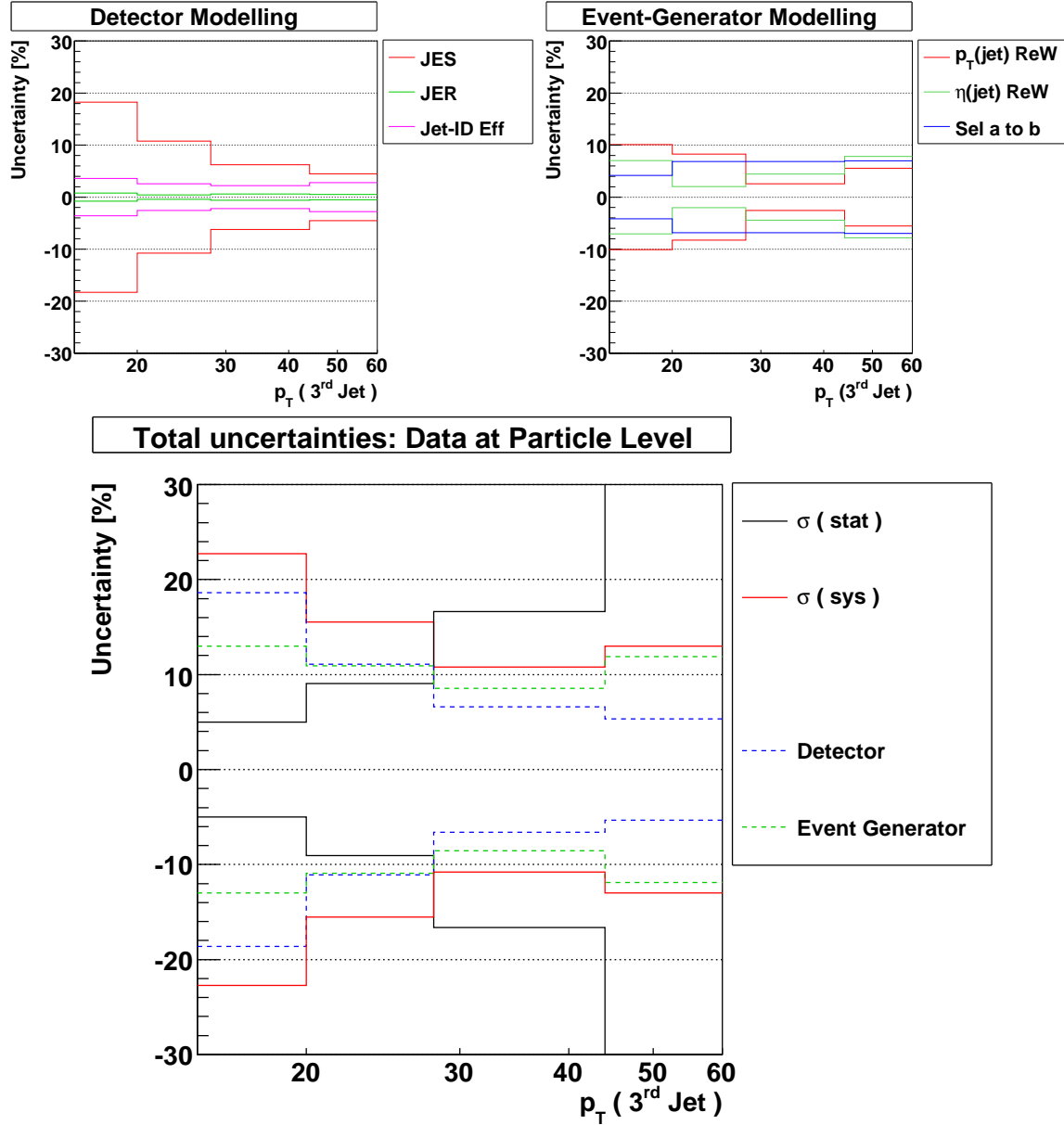


Figure 9.39: The systematic uncertainty of the measurement of p_T (third jet) is split in two groups of contributions arising from the detector simulation (upper left) and from the event-generator simulation (upper right). In the lower figure the total statistical and systematic uncertainties are given, together with the sum in quadrature of the systematic uncertainties arising from detector and event-generator related corrections. All p_T^{jet} values are in units of GeV.

	Total		Detector Corr			Kinematic Corr		
$p_T(\text{Jet})$	σ_{stat}	σ_{sys}	JES	JER	$\epsilon_{ID}(\text{jet})$	$p_T(\text{jet})$	$\eta(\text{jet})$	Sel $a \rightarrow b$
[GeV]	[%]							
20 - 28	1.6	7.0	5.7	0.2	1.1	0.7	0.4	3.8
28 - 40	2.1	4.6	3.0	0.1	1.0	0.6	2.5	2.1
40 - 54	3.1	3.9	2.8	0.2	0.9	1.1	0.7	2.1
54 - 73	4.0	4.1	2.7	0.1	0.7	1.0	1.3	2.4
73 - 95	5.6	8.2	3.9	0.1	0.5	4.9	4.2	3.1
95 - 120	8.4	8.9	4.4	0.3	0.8	5.7	4.1	3.1
120 - 154	12.0	10.9	5.9	0.1	1.2	7.0	4.7	3.2
154 - 200	17.1	14.2	12.3	0.2	0.4	3.4	5.4	3.3
200 - 300	41.9	25.2	20.6	0.1	1.1	7.7	11.7	3.6

Table 9.6: Overview of the different uncertainties of the measurement of p_T (leading jet). The main source of systematic uncertainty is the jet energy scale (JES) corrections. Other sources of systematic uncertainty is the jet energy resolution (JER), the jet identification efficiency ($\epsilon_{ID}(\text{jet})$), as well as the p_T^{jet} and η^{jet} reweighting. The extrapolation from selection a to selection b is an additional source of uncertainty for the results quoted without selection criteria on the electron p_T and y . The various systematic uncertainties were combined in quadrature to give the total systematic uncertainty.

	Total		Detector Corr			Kinematic Corr		
$p_T(\text{Jet})$	σ_{stat}	σ_{sys}	JES	JER	$\epsilon_{ID}(\text{jet})$	$p_T(\text{jet})$	$\eta(\text{jet})$	Sel $a \rightarrow b$
[GeV]	[%]							
20 - 28	3.7	10.9	9.7	0.3	1.7	1.3	2.2	3.9
28 - 40	5.7	6.0	4.4	0.3	1.0	2.3	1.0	2.9
40 - 54	9.1	6.8	4.1	0.9	2.5	3.3	1.8	3.0
54 - 73	15.3	9.0	5.7	0.9	2.8	2.8	4.8	3.0
73 - 200	22.1	8.0	4.0	0.0	1.4	5.3	3.0	3.1

Table 9.7: Overview of the different uncertainties of the measurement of p_T (second jet). See the caption of Table 9.6 for details about the labelling of the various sources of systematic uncertainty.

	Total		Detector Corr			Kinematic Corr		
$p_T(\text{Jet})$	σ_{stat}	σ_{sys}	JES	JER	$\epsilon_{ID}(\text{jet})$	$p_T(\text{jet})$	$\eta(\text{jet})$	Sel $a \rightarrow b$
[GeV]	[%]							
20 - 28	9.1	15.5	10.8	0.4	2.6	8.3	2.0	6.8
28 - 44	16.6	10.8	6.2	0.6	2.2	2.5	4.4	6.8
44 - 60	42.2	13.0	4.5	0.5	2.8	5.6	7.8	7.0

Table 9.8: Overview of the different uncertainties of the measurement of p_T (third jet). See the caption of Table 9.6 for details about the labelling of the various sources of systematic uncertainty.

			$65 < M_{ee} < 115 \text{ GeV}, p_T^e > 25 \text{ GeV},$ $ y^e < 1.1 \text{ or } 1.5 < y^e < 2.5$			$65 < M_{ee} < 115 \text{ GeV}$		
$p_T(\text{Jet})$ [GeV]	Bin ctr. [GeV]	# of Events	$\frac{1}{\sigma_{Z/\gamma^*}} \times \frac{d\sigma}{dp_T(\text{jet})}$ [1/GeV]	σ_{stat} [%]	σ_{sys} [%]	$\frac{1}{\sigma_{Z/\gamma^*}} \times \frac{d\sigma}{dp_T(\text{jet})}$ [1/GeV]	σ_{stat} [%]	σ_{sys} [%]
20 - 28	23.6	3795.6	6.81×10^{-3}	1.6	5.9	7.19×10^{-3}	1.6	7.0
28 - 40	33.4	2345.4	2.99×10^{-3}	2.1	4.0	3.22×10^{-3}	2.1	4.6
40 - 54	46.3	1077.6	1.23×10^{-3}	3.1	3.3	1.37×10^{-3}	3.1	3.9
54 - 73	62.4	634.7	5.04×10^{-4}	4.0	3.3	5.74×10^{-4}	4.0	4.1
73 - 95	82.6	318.8	2.03×10^{-4}	5.6	7.6	2.27×10^{-4}	5.6	8.2
95 - 120	105.8	143.6	7.29×10^{-5}	8.4	8.4	7.62×10^{-5}	8.4	8.9
120 - 154	133.9	70.9	2.64×10^{-5}	12.0	10.4	2.54×10^{-5}	12.0	10.9
154 - 200	171.7	34.7	8.08×10^{-6}	17.1	13.8	6.99×10^{-6}	17.1	14.2
200 - 300	229.2	5.8	7.46×10^{-7}	41.9	24.9	5.58×10^{-7}	41.9	25.2

Table 9.9: Summary of the $p_T(\text{leading jet})$ measurement. For each bin the following information is given: bin center; number of events after background subtraction; and the measured cross section with uncertainties for both selection a and b .

measured distributions of $\frac{1}{\sigma_{Z/\gamma^*}} \times \frac{d\sigma}{dp_T}$ for the N^{th} jet in $Z/\gamma^* \rightarrow e^+e^-$ events with at least N jets are summarized in Table 9.9 for $N = 1$, Tab. 9.10 for $N = 2$ and Tab. 9.11 for $N = 3$. Each measurement is specified both for selection a and selection b and is normalized per 1 GeV. For each bin the *corrected bin center* is specified, following the method given in Ref. [126]. Given a continuous prediction of an observable $f(x)$ which is then binned in x and normalized per x unit, the corrected center of each bin is the value x_1 at which $f(x_1)$ is equal to the bin content. The corrected bin centers were derived using a LO MCFM prediction for each observable using a histogram with 1 GeV binning. The continuous prediction was estimated as the curve given by a linear extrapolation between the bin centers of the MCFM prediction. The sensitivity of the corrected bin center p_T value to the slope of the distribution was found to be negligible by reweighting the MCFM spectrum to decrease by a factor of ten more or less than data when going from 20 GeV to 300 GeV.

The measured cross sections per bin for selection b vary over three orders of magnitude from $5.8 \cdot 10^{-2} \times \sigma_{Z/\gamma^*}$ or 15 pb in the 20–28 GeV bin for $p_T(\text{leading jet})$, to $5.6 \times 10^{-5} \times \sigma_{Z/\gamma^*}$, or 15 fb. As a comparison, the latter number is about 70 times smaller than the cross section for Standard Model Higgs production at the Tevatron assuming $M_H = 120 \text{ GeV}$ [127].

For $p_T(\text{leading jet})$ in the range $28 < p_T < 73 \text{ GeV}$ the measurement has a total uncertainty, defined as the statistical and systematic uncertainty added in quadrature, of less than 6%. Since the uncertainty of the integrated luminosity of the DØ experiment is 6.1%, this means that the differential, normalized $p_T(\text{leading jet})$ measurement has a higher precision in this range than is possible to achieve for in a DØ measurement of the inclusive $Z/\gamma^* \rightarrow e^+e^-$ cross section itself.

9.21 Comparisons with existing measurements

As a cross-check, the measurements presented above are compared with two previously published $Z/\gamma^* + \text{jets}$ measurements from the DØ experiment: a jet multiplicity measurement [128] in the electron channel analysing a data set corresponding to $\mathcal{L} = 0.4 \text{ fb}^{-1}$; and a measurement of $p_T(\text{leading jet})$ in the muon channel [129] using $\mathcal{L} = 1.0 \text{ fb}^{-1}$.

			65 < M_{ee} < 115 GeV, $p_T^e > 25$ GeV, $ y^e < 1.1$ or $1.5 < y^e < 2.5$			65 < M_{ee} < 115 GeV		
$p_T(\text{Jet})$ [GeV]	Bin ctr. [GeV]	# of Events	$\frac{1}{\sigma_{Z/\gamma^*}} \times \frac{d\sigma}{dp_T(\text{jet})}$ [1/GeV]	σ_{stat} [%]	σ_{sys} [%]	$\frac{1}{\sigma_{Z/\gamma^*}} \times \frac{d\sigma}{dp_T(\text{jet})}$ [1/GeV]	σ_{stat} [%]	σ_{sys} [%]
20 - 28	23.6	725.4	1.30×10^{-3}	3.7	10.2	1.39×10^{-3}	3.7	10.9
28 - 40	33.2	310.1	4.23×10^{-4}	5.7	5.2	4.51×10^{-4}	5.7	6.0
40 - 54	46.1	122.2	1.57×10^{-4}	9.1	6.2	1.62×10^{-4}	9.1	6.8
54 - 73	61.9	43.8	4.17×10^{-5}	15.3	8.5	4.20×10^{-5}	15.3	9.0
73 - 200	98.1	21.7	2.96×10^{-6}	22.1	7.4	2.82×10^{-6}	22.1	8.0

Table 9.10: Summary of the p_T (second jet) measurement. For each bin the following information is given: bin center; number of events after background subtraction; and the measured cross section with uncertainties for both selection a and b .

			65 < M_{ee} < 115 GeV, $p_T^e > 25$ GeV, $ y^e < 1.1$ or $1.5 < y^e < 2.5$			65 < M_{ee} < 115 GeV		
$p_T(\text{Jet})$ [GeV]	Bin ctr. [GeV]	# of Events	$\frac{1}{\sigma_{Z/\gamma^*}} \times \frac{d\sigma}{dp_T(\text{jet})}$ [1/GeV]	σ_{stat} [%]	σ_{sys} [%]	$\frac{1}{\sigma_{Z/\gamma^*}} \times \frac{d\sigma}{dp_T(\text{jet})}$ [1/GeV]	σ_{stat} [%]	σ_{sys} [%]
20 - 28	23.5	122.8	2.22×10^{-4}	9.1	14.0	2.33×10^{-4}	9.1	15.5
28 - 44	34.2	36.6	4.40×10^{-5}	16.6	8.4	4.48×10^{-5}	16.6	10.8
44 - 60	50.2	5.8	8.67×10^{-6}	42.2	11.0	8.60×10^{-6}	42.2	13.0

Table 9.11: Summary of the p_T (third jet) measurement. For each bin the following information is given: bin center; number of events after background subtraction; and the measured cross section with uncertainties for both selection a and b .

The jet multiplicity measurement corrects data to the particle level with the selection criteria $75 < M_{ee} < 105$ GeV. Jets are defined using $R = 0.5$ DØ Run II mid-point jets satisfying $p_T^{\text{jet}} > 20$ GeV and $|\eta| < 2.5$. The measurement is normalized to the inclusive $Z/\gamma^* \rightarrow e^+e^-$ cross section and multiplied with 10^3 . The present measurements were integrated in p_T^{jet} to find the jet multiplicities and the two measurements are compared in Tab. 9.12. The two

Jet multiplicity	Previous				Present			
	R	σ_{stat}	σ_{sys}	$\sigma_{stat \oplus sys}$	R	σ_{stat}	σ_{sys}	$\sigma_{stat \oplus sys}$
≥ 0	1000	0	0	0	1000	0	0	0
≥ 1	120.1	3.3	16	16.3	134.4	1.4	5.6	5.8
≥ 2	18.6	1.4	6	6.2	19.9	0.6	1.3	1.4
≥ 3	2.8	0.56	1.4	2.9	2.7	0.2	0.3	0.4

Table 9.12: Comparison of the inclusive jet multiplicities, normalized to the inclusive sample and multiplied by 10^3 , between the present result and a previously published DØ result [128].

measurements are found to be in agreement within the quoted uncertainties. The systematic uncertainties of the present analysis are reduced by about a factor of three compared with the earlier result. The reduction in statistical uncertainty is larger than expected from the increase in luminosity, indicating that the $(\mathcal{A} \times \epsilon)$ of the present analysis is higher than for the earlier result.

Next, comparisons with the measurement of p_T (leading jet) in the $Z/\gamma^* \rightarrow \mu\mu$ channel are

shown. Assuming lepton universality the two measurements are compared by extrapolating the present result to the binning and particle-level selection used in the muon-channel result, namely $|y^\mu| < 1.7$ and $65 < M_{\mu\mu} < 115$ GeV. The p_T spectrum from the muon analysis was normalized to the Z/γ^* cross section of 118 pb which is quoted in [129]. The ratio of the muon-channel result to the present result is given in Fig. 9.40. Only statistical uncertainties

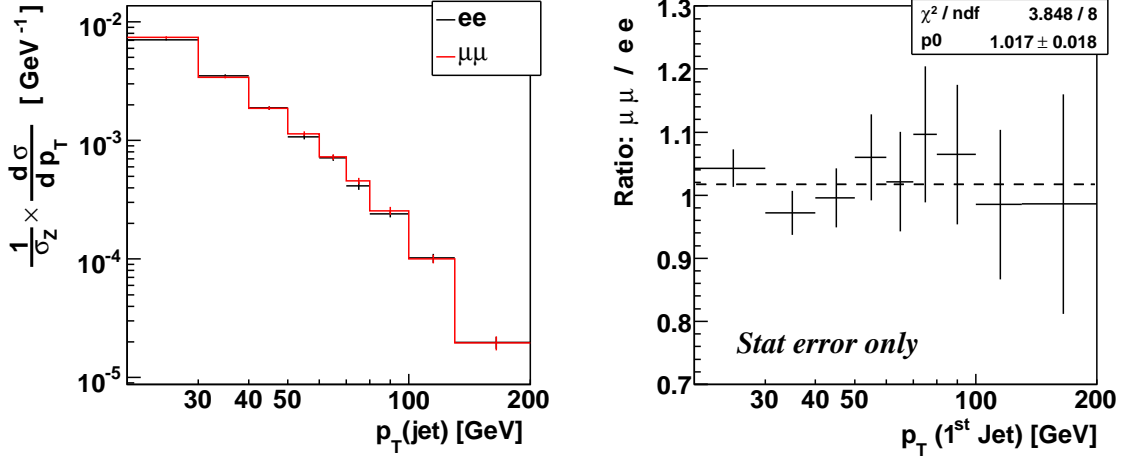


Figure 9.40: Ratio of the muon-channel measurement of $p_T(\text{leading jet})$ to the present result. The ratio was fitted with a constant, testing the hypothesis that the two measurements are compatible, giving 1.017 ± 0.018 with a χ^2/DOF of 3.8/8.

are taken into account since most of the systematic uncertainties are fully correlated between the two measurements. The ratio between the two results was fitted with a constant, testing the hypothesis that the two measurements are compatible, giving 1.017 ± 0.018 with a χ^2/dof of ⁵ 3.8/8 indicating that the two measurements agree.

Direct comparisons have not been performed with a CDF measurement [130] of the inclusive p_T^{jet} spectra⁶ in $Z/\gamma^* \rightarrow e^+e^-$ events with N or more jets ($N = 1, 2$) using a data set corresponding to $\mathcal{L} = 1.7 \text{ fb}^{-1}$. In Ref. [130] the CDF result is shown to be in good agreement with a fixed-order NLO pQCD calculation from MCFM, and the same is true for the present measurement as is shown in the next Chapter. Based on this one can indirectly conclude that the two measurements are in agreement. A direct comparison has been performed between the uncertainties of the two measurements using CDF numbers for $p_T(N^{\text{th}} \text{ jet})$ in events with at least N jets, $N = 1, 2$, which are available in Ref. [131]. For the comparison it is a problem that the CDF measurement uses a finer binning than the present measurement. After rebinning, the comparison between the respective statistical and systematic uncertainties are given in Fig. 9.41. It is seen that the statistical uncertainties of the two measurement are comparable, in spite of the 1.7 times larger data set analysed in the CDF study. The systematic uncertainties of the present measurements are smaller than in the CDF study in most bins. For $p_T(\text{leading jet})$ below 70 GeV, where both measurements are systematically limited, the systematic uncertainties of the CDF measurement are 50–90% larger than in the

⁵The muon result was corrected to particle level using a regularized unfolding program called GURU [107].

This algorithm smooths the ratio of data to simulation, thereby removing fluctuations in the data spectrum.

This effect was not taken into account when computing the χ^2/dof .

⁶In this context *inclusive* means that the p_T of each jet in the event is entered in the same histogram.

present analysis. For the second jet the systematic uncertainties of the CDF measurement are 30–40 larger than in the present analysis while the statistical uncertainties are of comparable size.

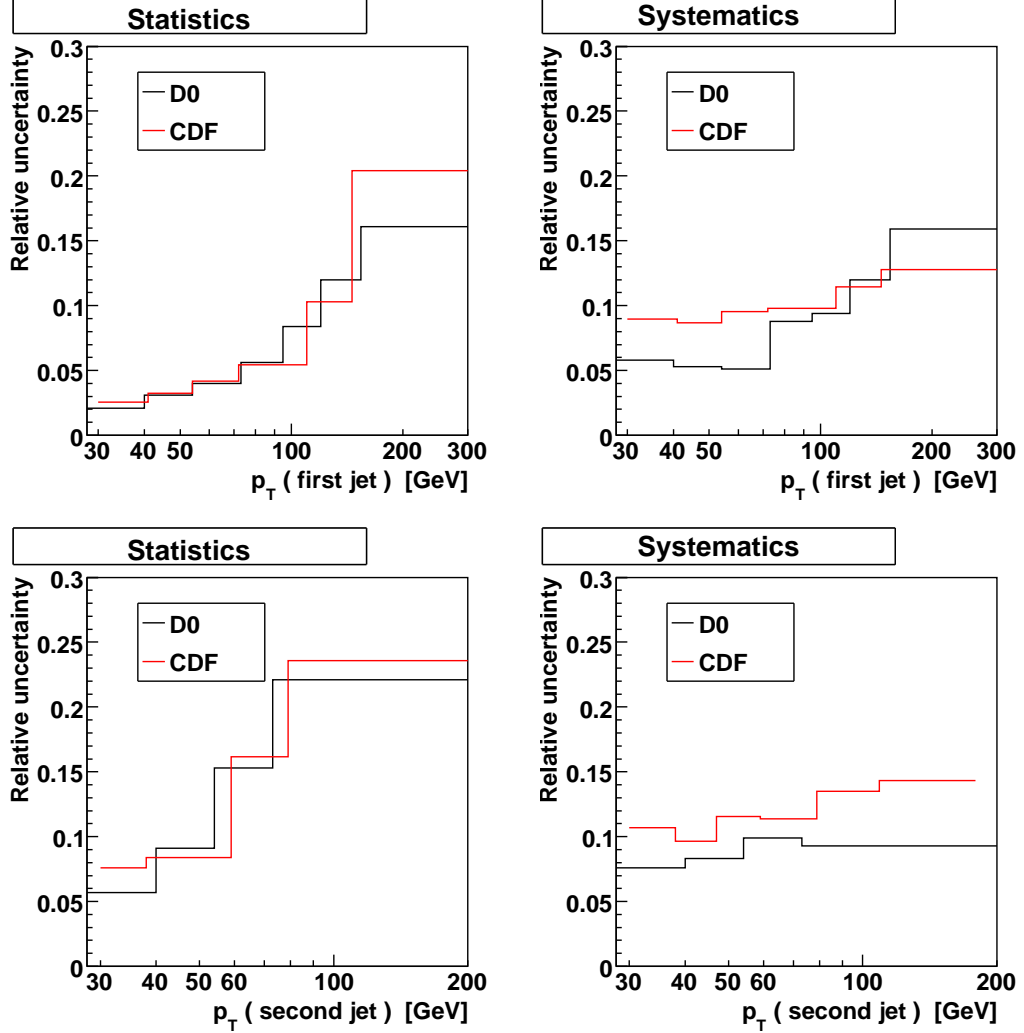


Figure 9.41: The relative statistical and systematic uncertainties of the CDF measurement [130] compared with those of the measurements presented in this Chapter, for $p_T(\text{leading jet})$ (above) and $p_T(\text{second jet})$. While the statistic uncertainties are of comparable size, the present measurement generally has lower systematic uncertainties.

9.22 Measurement of the Differential $\Delta\phi(\text{di-electron, jet})$ Distribution

It was seen above that ALPGEN+PYTHIA does not describe the distribution of $\Delta\phi(\text{di-electron, leading jet})$ observed in data. By reweighting ALPGEN+PYTHIA to SHERPA at the particle

level two-dimensionally in p_T^Z and $\Delta\phi(Z, \text{leading jet})$ an improved level of agreement with data was achieved. The measurement of $\Delta\phi(\text{di-electron, jet})$ presented below, corrected to the particle level, can serve as a benchmark for improving the ALPGEN+PYTHIA model as well as for a direct comparison of the SHERPA model with data.

The $\Delta\phi(\text{di-electron, jet})$ measurement is normalized to unity since the jet production rates were already probed through the p_T^{jet} measurements given above. The shape of the distribution has significantly lower systematic uncertainties than the distribution itself. The measurement is presented for the particle-level electron selection $p_{T1} > 25$ GeV, $|y^e| < 1.1$ or $1.5 < |y^e| < 2.5$ and $65 < M_{ee} < 115$ GeV. Jets are required to satisfy $p_T > 20$ GeV and $\Delta R(e, \text{jet}) > 0.8$.

The migration matrix between $\Delta\phi(\text{di-electron, leading jet})$ at particle level ($\Delta\phi^{\text{ptcl}}$) and at detector level ($\Delta\phi^{\text{reco}}$), as predicted by ALPGEN+PYTHIA, is given in Fig. 9.42 (upper, left). A significant amount of off-diagonal elements are seen, with a local maximum corresponding to events where the particle-level $\Delta\phi^{\text{ptcl}}$ is close to π whereas the reconstructed $\Delta\phi^{\text{reco}}$ is ≈ 0 . The distribution of $\Delta\phi^{\text{ptcl}}$, for events corresponding to four selected bins in $\Delta\phi^{\text{reco}}$, is given in Fig. 9.42 (upper, right). For events with $\Delta\phi^{\text{reco}} \approx \pi$, $\Delta\phi^{\text{reco}} \approx \Delta\phi^{\text{ptcl}}$ for about 60% of the events. For events with $\Delta\phi^{\text{reco}} \approx 0.25$ the particle-level and detector-level $\Delta\phi$ are in the same bin in less than 20% of the cases, and the distribution has a local maximum at π , corresponding to the off-diagonal entries seen in the migration matrix. For events with $\Delta\phi^{\text{reco}} < 1.0$, the probability (*matching efficiency*) that the observable is defined at the particle level, i.e. that a particle-level jet exists, is only 30%, as illustrated in Fig. 9.42 (lower, left). As a consequence of this, the $(\mathcal{A} \times \epsilon)$ becomes large for small values of $\Delta\phi^{\text{reco}}$.

The large amount of off-diagonal elements in the migration matrix has been traced back to two classes of events: first, events where the direction of flight of the reconstructed di-electron system does not correspond to that of the particle-level di-electron system. For small values of p_T^Z the two Z -decay electrons have approximately equal p_T and the direction of flight of the reconstructed di-electron system is highly sensitive to the electron energy resolution. The distribution of $\Delta\phi$ between the di-electron systems, at the two levels, for various lower selection criteria on the detector-level p_T^Z , is given in Fig. 9.43. For inclusive events the reconstructed and true di-electron azimuthal angles deviate by more than one radian in $\sim 10\%$ of the events, whereas this number is reduced to $\sim 1\%$ for $p_T^Z > 5$ GeV. The latter selection criterion is adopted in this analysis to guarantee that the direction of flight of the di-electron system is well defined. The second class of events contributing to the off-diagonal entries in the migration matrix in Fig. 9.42 arises from the impact of the jet energy resolution on events which contain two particle-level jets with similar values of p_T . As seen in Fig. 9.33 (upper, center), this is the case for a large fraction of the events containing two or more jets. Due to the jet energy resolution there is a significant probability that the leading jet at particle level gives rise to the second leading jet at detector level. In the case of the p_T^{jet} measurements this gives rise to nearly diagonal entries in the migration matrix since the reconstructed p_T^{jet} values of the two jets are similar. For the $\Delta\phi$ measurement, however, the arising entries are in general off-diagonal. To make the measurement less sensitive to the impact of the jet energy resolution the observable is redefined as being inclusive in jets. More specifically, in each event the value of $\Delta\phi(\text{di-electron, jet})$ for every jet is taken into account, not only for the leading jet. The corresponding migration matrix, for events with $p_T^Z > 5$ GeV, is given in Fig. 9.44 (left). The lower p_T^Z criterion combined with evaluating $\Delta\phi$ for all jets in each event leads to a large reduction in the off-diagonal entries, and correspondingly to a more well-behaved observable. The particle-level distribution, for each detector-level bin, is peaked at the bin itself (center).

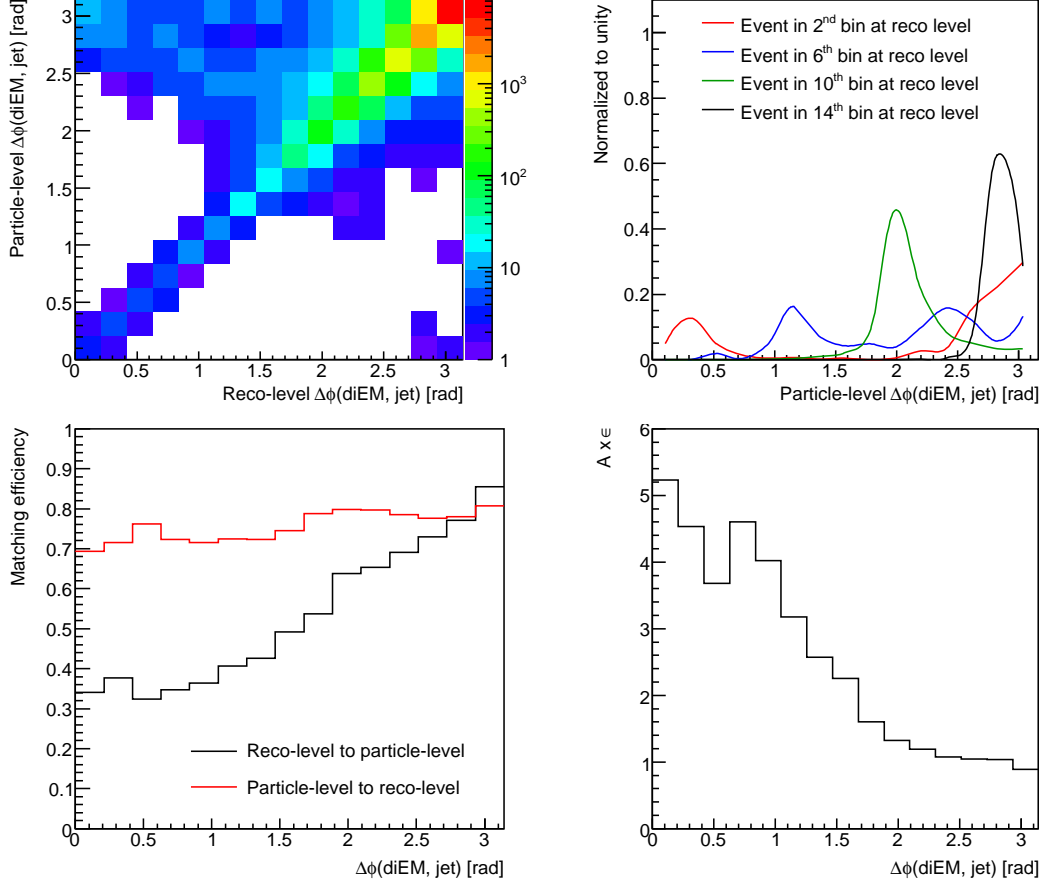


Figure 9.42: Large off-diagonal elements are seen in the simulated migration matrix between $\Delta\phi(\text{di-electron, leading jet})$ at the particle level and at the detector level (upper, left). An alternative presentation of the information contained in the migration matrix is the distribution of $\Delta\phi$ at the particle level for those events which contribute to a specific bin at the detector level (upper right). The lower, left figure shows the probability (*matching efficiency*) that an event which contributes to the observable at detector level also contributes at the particle level (labeled *Reco to particle*), and vice versa (labeled *Reco to particle*). It is seen that only 1/3 of the events which has $\Delta\phi < 1$ at the detector level has a jet at the particle level. The $(\mathcal{A} \times \epsilon)$ (lower, left) is therefore large for small values of $\Delta\phi$.

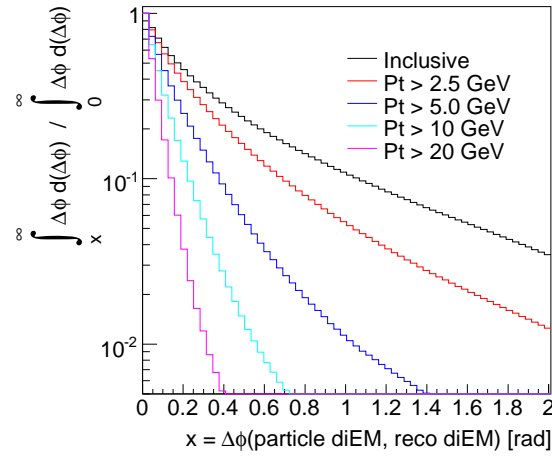


Figure 9.43: For each x and lower p_T^Z selection criterion (see legend), the probability that the difference in azimuthal angle between the particle-level and the detector-level di-electron system (labeled $\Delta\phi(\text{particle diEM, reco diEM})$) in a random event is larger than the value x . For example, in the case of no p_T^Z selection criterion (labeled *Inclusive*), $\Delta\phi$ exceeds 1.0 radian in about 10% of the events. For $p_T^Z > 5$ GeV the corresponding number is about 1%. This indicates that the direction of flight of the di-electron system is highly sensitive to the electron p_T resolution for low values of p_T^Z and its azimuthal angle is therefore not measurable for such events. By requiring $p_T^Z > 5$ GeV the azimuthal angle of the di-electron system is reasonably well measured.

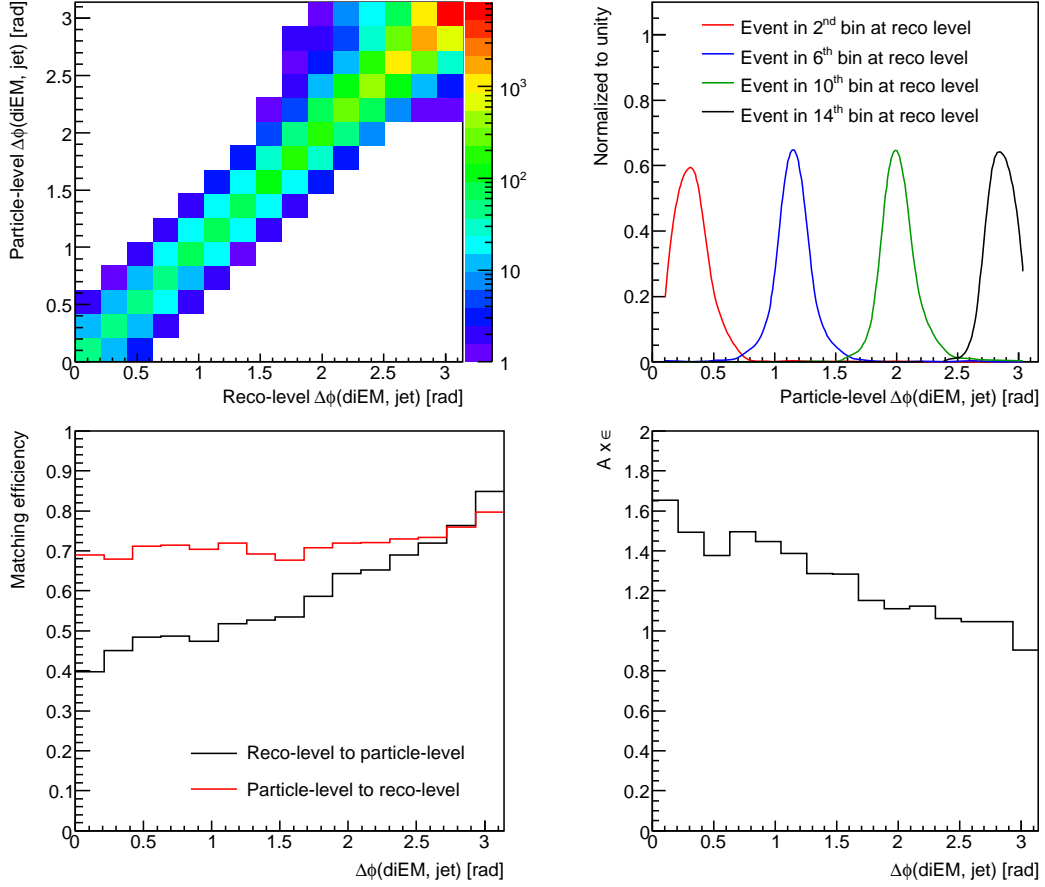


Figure 9.44: Same distributions as in Fig. 9.42, but for the $\Delta\phi$ observable being evaluated for all (di-electron, jet) pairs in each event. The reconstructed di-electron p_T was required to exceed 5 GeV.

The event sample still contains a large number of events with a detector-level $\Delta\phi$ value below ~ 1 where the observable is not defined at the particle level, as seen from the low matching efficiency for detector-level events. Correspondingly, the $(\mathcal{A} \times \epsilon)$ is large for this region of phase space. To determine where the jets giving rise to the detector-level observable in these events come from, detector-level jets were matched in ΔR with particle-level particles. A jet is said to have particle-level confirmation if the sum of the p_T of all particles satisfying $\Delta R(\text{particle, jet axis}) < 0.5$ exceeds $1/4$ of the value of p_T^{jet} . The $\Delta\phi$ distribution of detector-level jets which do not have particle-level confirmation contains a class of jets which are constant in $\Delta\phi$, see Fig. 9.45, compatible with jets arising from pile-up $p\bar{p}$ interactions. Reconstructed jet candidates from pile-up interactions are seen to constitute a sizable background to the measurement for small values of $\Delta\phi$. To reject jet candidates from pile-up interactions the difference in z coordinate between the primary interaction vertex (PV) and any vertex from additional $p\bar{p}$ interactions was exploited. Reconstructed track candidates were selected using the selection criteria described in Sect. 9.13.2. A jet is matched to a track if $\Delta R(\text{jet-axis,}$

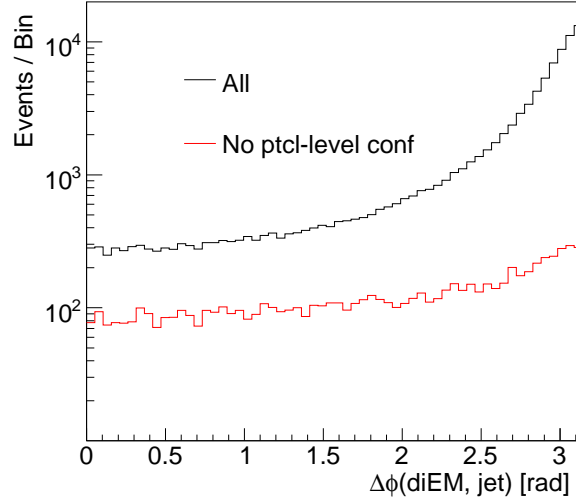


Figure 9.45: The $\Delta\phi$ distribution for all reconstructed jets and for those reconstructed jets for which the p_T -sum of all generated particles which satisfy $\Delta R(\text{particle, jet-axis}) < 0.5$ is less than $1/4$ of the reconstructed p_T^{jet} value (labeled *No ptcl-level conf*).

track) < 0.5 . Tracks having a difference in z coordinate of less than 2 cm with respect to the PV⁷ are labeled as *PV tracks*, all other tracks as *non-PV tracks*. It was validated on an event-by-event basis that jets without particle-level confirmation tend to be matched with tracks pointing to a different interaction vertex than the PV. The PV track- p_T fraction of a jet is defined as the ratio of the p_T sum of all PV tracks matched to the jet ($\sum p_{T,\text{track}}^{\text{PV}}$) to the p_T sum of all tracks matched to the jet ($\sum p_{T,\text{track}}^{\text{all}}$). In Fig. 9.46 the distribution of $\sum p_{T,\text{track}}^{\text{PV}} / \sum p_{T,\text{track}}^{\text{all}}$ is shown as function of $\sum p_{T,\text{track}}^{\text{all}}$ for simulated jets without (left) and with (right) particle-level confirmation. To reject jets from pile-up interactions the selection criteria $\sum p_{T,\text{track}}^{\text{PV}} / \sum p_{T,\text{track}}^{\text{all}} > 0.7$ and $\sum p_{T,\text{track}}^{\text{all}} > 5$ GeV were applied. The selection efficiency is $\sim 90\%$ for simulated jets with particle-level confirmation against $\sim 2\%$ for jets lacking such confirmation. The application of this anti pile-up jet selection removes the class of jets which lacks particle-level confirmation and which is independent of $\Delta\phi$, as seen in Fig. 9.47.

The migration matrix for events with $p_T^Z > 5$ GeV and only jets passing the anti-pile-up selection being taken into account is given in Fig. 9.44. It is seen that the removal of jets which are not compatible with arising from the primary interaction vertex leads to a matching efficiency for detector-level events which is above 0.6 for all values of $\Delta\phi$. The $(\mathcal{A} \times \epsilon)$ which is close to unity, meaning that the observable is almost insensitive to detector effects.

To determine the dependency of the $(\mathcal{A} \times \epsilon)$ for $\Delta\phi(\text{di-electron, jet})$ on the event generator model, this factor is evaluated for the ALPGEN+PYTHIA once with no kinematic reweighting, and once with the two-dimensional p_T^Z and $\Delta\phi(Z, \text{leading jet})$ reweighting to SHERPA and then one-dimensional p_T^Z reweighting to data. The corresponding level of agreement with data at the detector level is shown in Fig. 9.49. The central prediction of $(\mathcal{A} \times \epsilon)$ is evaluated with the kinematic reweighting applied. The difference between applying and not applying

⁷For comparison, the resolution of the z -component of the primary vertex reconstruction is about $35\mu\text{m}$.

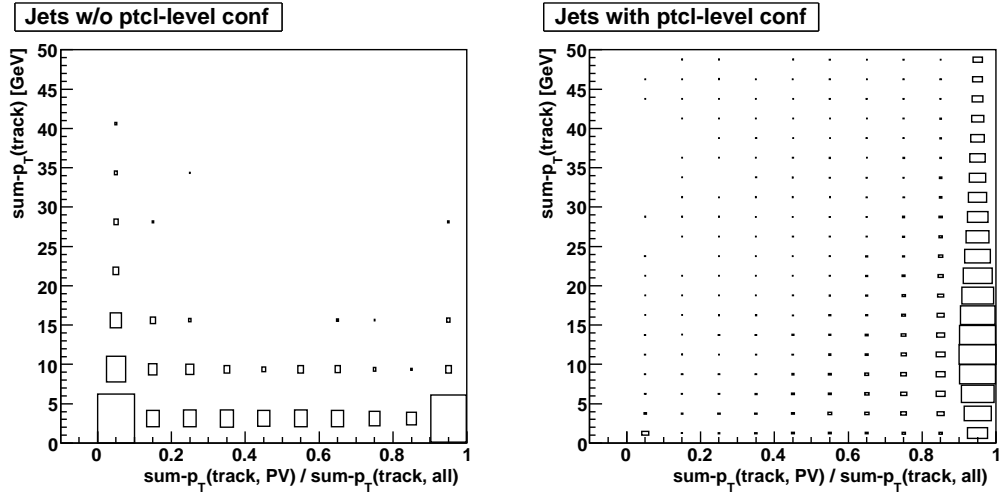


Figure 9.46: The correlation between $\sum p_{T,track}^{all}$ and $\sum p_{T,track}^{PV} / \sum p_{T,track}^{all}$ for jets without (left) and with (right) particle-level confirmation.

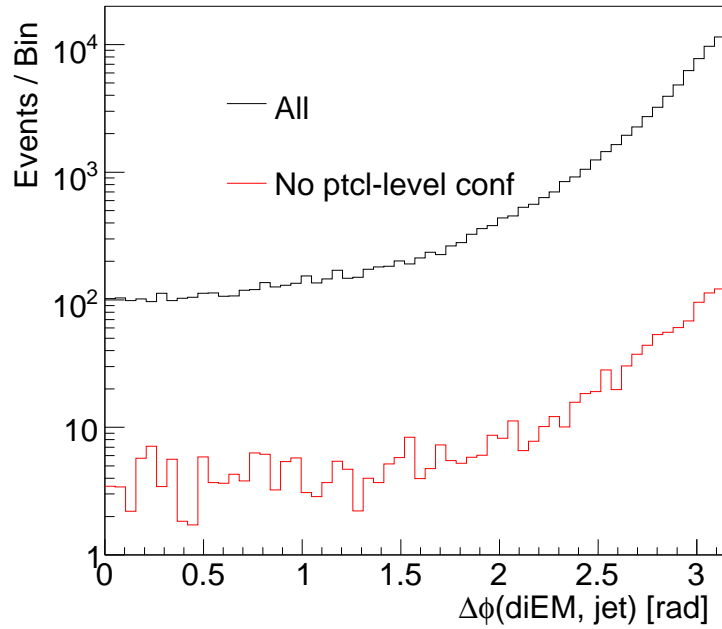


Figure 9.47: The $\Delta\phi$ distribution for all reconstructed jets and for those without particle-level confirmation after the application of the anti pile-up jet selection.

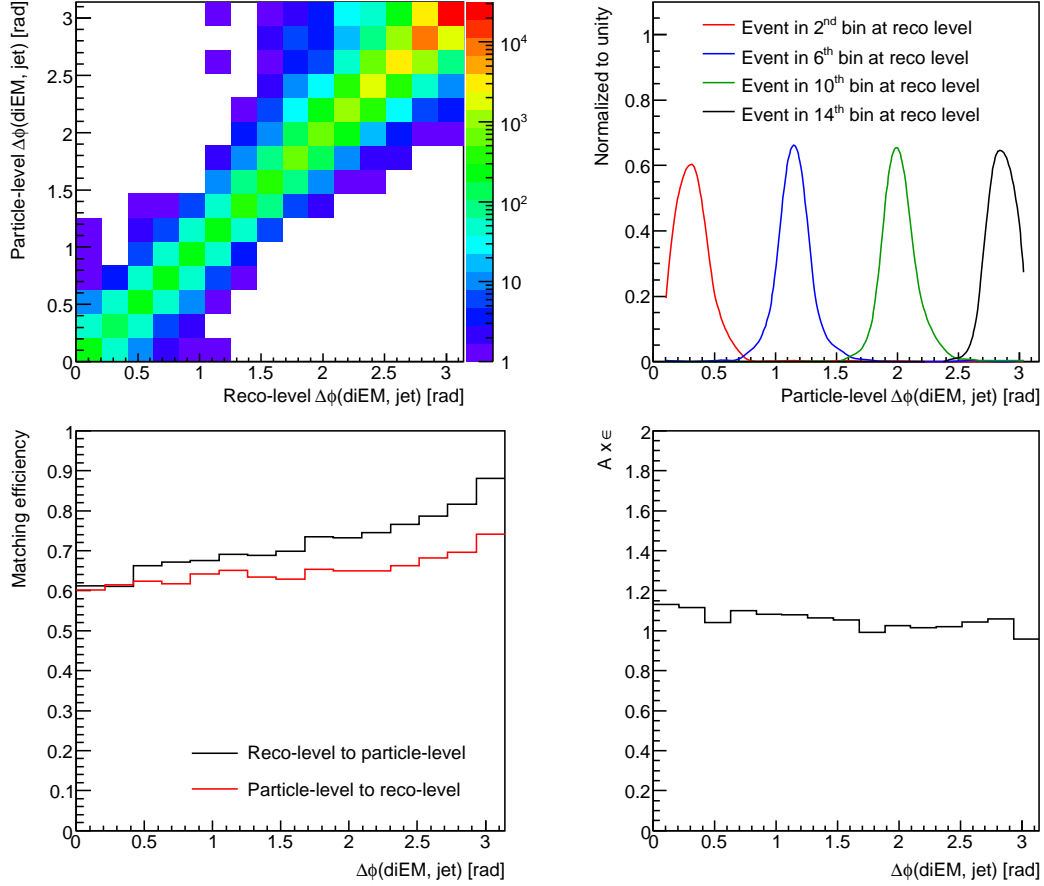


Figure 9.48: Same distributions as in Fig. 9.42, but for the $\Delta\phi$ observable being evaluated for all (di-electron, jet) pairs in each event, and only jets which are compatible with stemming from the primary interaction taken into account. The reconstructed di-electron p_T was required to exceed 5 GeV. The particle-level and detector-level observable is seen to be highly correlated. In addition, the matching efficiency is of comparable size both for detector-level and particle-level events.

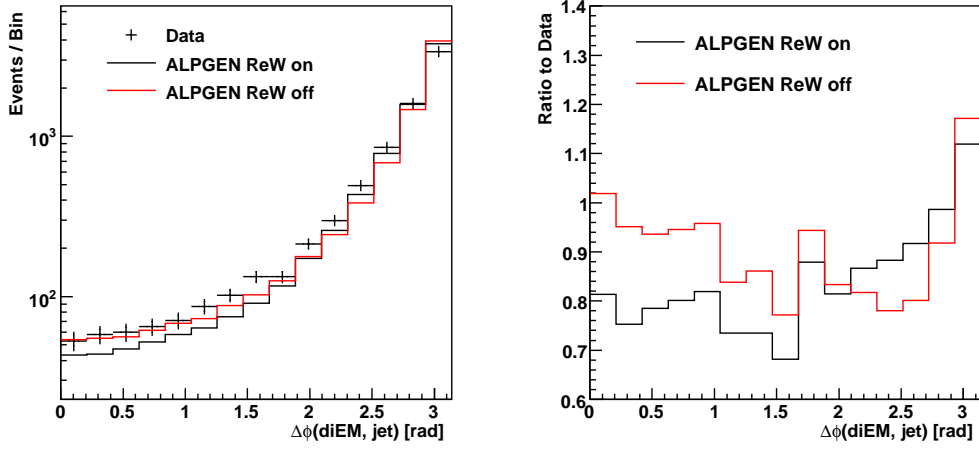


Figure 9.49: ALPGEN+PYTHIA with and without kinematic reweighting compared with data.

the reweighting is assigned as a systematic uncertainty of the measurement. Additional sources of systematic uncertainty to the factor $(\mathcal{A} \times \epsilon)$ arise from varying the jet energy scale and resolution and jet identification efficiency by $\pm 1\sigma$. The resulting variations in $(\mathcal{A} \times \epsilon)$ for $\Delta\phi(\text{di-electron, jet})$ are shown in Fig. 9.50. The total upwards (downwards) systematic uncertainty of the measurement is defined as the sum in quadrature of the positive (negative) variations for each bin. The measurement is summarized in Tab. 9.13, with the integral of the $\Delta\phi(\text{di-electron, jet})$ distribution being normalized to unity. This choice of normalization significantly reduces the sensitivity of the measurement to the uncertainty of the jet energy scale. The measurement is statistically limited in all bins of the measurement. In the next Chapter the measurement is compared with the predictions of various commonly used event generator models.

Bin Range	$\sigma^{-1} d\sigma/d\Delta\phi$ Radian $^{-1}$	σ_{stat} %	σ_{sys}^{up} %	σ_{sys}^{down} %	σ_{ReW} %	σ_{JES}^{Up} %	σ_{JES}^{Down} %	σ_{JER}^{Up} %	σ_{JER}^{Down} %	σ_{ID}^{Up} %	σ_{ID}^{Down} %
0 – $\pi/15$	2.71×10^{-2}	14.7	1.4	-8.0	-7.5	1.3	-2.9	0.3	-0.1	-0.4	0.4
$\pi/15$ – $2\pi/15$	3.30×10^{-2}	13.9	4.8	-2.7	3.3	3.5	-2.7	0.2	0.1	-0.3	0.3
$2\pi/15$ – $3\pi/15$	3.64×10^{-2}	13.7	3.8	-1.4	0.1	3.7	-1.4	0.5	0.4	-0.4	0.4
$3\pi/15$ – $4\pi/15$	3.77×10^{-2}	13.2	4.7	-2.8	4.3	1.8	-2.8	0.2	-0.2	0.0	-0.0
$4\pi/15$ – $5\pi/15$	4.26×10^{-2}	12.6	6.3	-1.7	6.0	2.0	-1.7	0.4	0.1	-0.2	0.2
$5\pi/15$ – $6\pi/15$	5.10×10^{-2}	11.5	3.1	-1.6	2.2	2.1	-1.6	-0.1	-0.3	-0.3	0.3
$6\pi/15$ – $7\pi/15$	5.94×10^{-2}	10.5	3.8	-2.0	2.2	2.9	-2.0	0.8	0.5	-0.4	0.4
$7\pi/15$ – $8\pi/15$	7.56×10^{-2}	9.2	1.2	-3.3	-2.3	1.1	-2.4	0.1	0.1	-0.5	0.5
$8\pi/15$ – $9\pi/15$	8.36×10^{-2}	9.1	3.0	-1.8	2.4	1.7	-1.8	0.2	-0.0	-0.2	0.2
$9\pi/15$ – $10\pi/15$	1.29×10^{-1}	7.3	3.3	-1.2	0.9	3.0	-1.2	0.8	0.6	-0.7	0.7
$10\pi/15$ – $11\pi/15$	1.85×10^{-1}	6.2	2.2	-1.7	1.5	1.6	-1.7	0.2	0.0	-0.2	0.2
$11\pi/15$ – $12\pi/15$	3.00×10^{-1}	4.8	1.2	-1.5	1.1	0.5	-1.3	-0.1	-0.1	0.6	-0.6
$12\pi/15$ – $13\pi/15$	5.21×10^{-1}	3.7	3.9	-0.6	3.7	1.1	-0.6	0.2	0.1	-0.1	0.1
$13\pi/15$ – $14\pi/15$	9.44×10^{-1}	2.7	3.1	-0.4	3.1	-0.1	-0.3	-0.1	-0.1	0.3	-0.3
$14\pi/15$ – π	2.22×10^0	1.9	0.9	-1.8	-1.6	-0.8	0.9	-0.1	0.0	-0.1	0.1

Table 9.13: Summary of the measurement of the shape of the jet-inclusive $\Delta\phi(\text{di-electron, jet})$ distribution for $p_T^Z > 5$ GeV. The table the various systematic uncertainties arising from the kinematic reweighting of the simulated event sample (σ_{ReW}), from the jet energy scale (σ_{JES}), from the jet energy resolution (σ_{JER}) and from the jet identification efficiency (σ_{ID}).

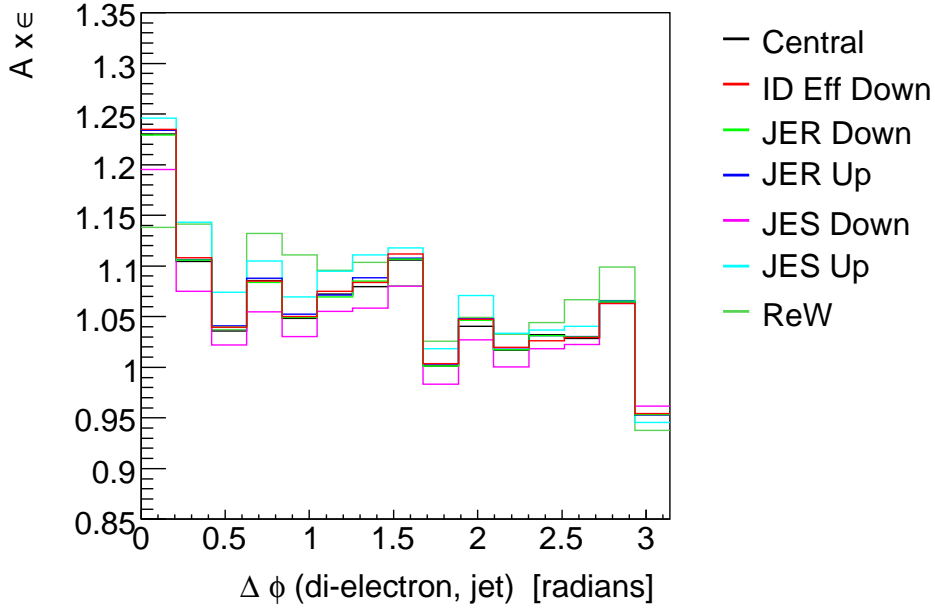


Figure 9.50: The $(\mathcal{A} \times \epsilon)$ for the jet-inclusive $\Delta\phi(\text{di-electron, jet})$ distribution. The distribution is given for simulation with kinematic reweighting (labeled *Central*) and with no such reweighting (*ReW*). In addition, the $(\mathcal{A} \times \epsilon)$ is given for the jet energy scale and resolution varied by $\pm 1\sigma$, and for the jet identification efficiency varied by -1σ .

Chapter 10

Comparing Theory Predictions with the Measurements

Below, the measurements presented in the previous chapter are compared with the predictions of the various theory models described in Chapter 3. First, the predictions of fixed-order calculations at leading and next-to-leading order in α_s are compared with the p_T^{jet} measurements as a test of perturbative QCD. Next, the p_T^{jet} and $\Delta\phi(\text{di-electron, jet})$ measurements are used as the basis for a detailed, comparative study of the particle-level predictions of the SHERPA, ALPGEN+PYTHIA, ALPGEN+HERWIG, MC@NLO, PYTHIA and HERWIG event-generator models. In addition, the discrepancies between PYTHIA and data for the distributions of $\Delta\phi(\text{jet, jet})$ seen in the detector-level study presented in Chapter 7 are interpreted using particle-level comparisons between the various event-generator models.

10.1 Theory Predictions

The theoretical aspects of the various simulation codes were described in detail in Chapter 3; only technique details are covered here. All theory predictions were generated using commonly available codes provided by the authors of the respective tools. The PDFs used were the CTEQ 6.1M [49] set¹, accessed through the LHAPDF interface [132]. The evolution of α_s was performed using the NLO evolution equation.

10.1.1 SHERPA

The SHERPA samples were generated using version 1.1.1 of the code. Up to three QCD partons were included in the matrix-element calculations. The configuration parameters were adopted from the example configuration file in the directory in the SHERPA release called *Tevatron1800*. The matching between the matrix-element and parton-shower parts of the algorithm was defined by setting the parameters of the k_T jet algorithm to $p_T^{\text{cut}} = 15$ GeV and $D = 1.0$ (see Sect. 3.7.2). The default choices of μ_F and μ_R were taken to be those defined by the CKKW algorithm. A patch [104] was applied to the code to correct for a bug which caused variations in μ_F to not be propagated properly to the LHAPDF library.

¹This set of PDFs is defined using NLO cross section calculations and the NLO evolution of α_s . For consistency, the LO MCFM calculation presented below should be made using a LO PDF set. For this calculation it was tested to use the LO CTEQ 6LL set together with the LO evolution of α_s , but only minor differences were seen with respect to using CTEQ 6.1M and the NLO α_s . Below, the latter PDF is used for all predictions.

10.1.2 ALPGEN+PYTHIA and ALPGEN+HERWIG

Release version v2.13 of ALPGEN was used to generate samples using either PYTHIA v6.325 using tune QW [133] or HERWIG v6.510 + JIMMY v4.31 to provide the parton shower algorithm and modeling of hadronization and the underlying event. For JIMMY the default configuration parameters were used. Up to four QCD partons were included in the matrix-element calculations, and the matching between matrix element and parton shower partons were defined by the cuts $p_T > 15$ GeV and $\Delta R = 0.4$. The nominal value of μ_F was taken to be $\mu_F = \sqrt{M_Z^2 + p_{T,Z}^2}$, with M_Z and $p_{T,Z}^Z$ denoting the mass and transverse momentum of the Z/γ^* boson. The nominal value of μ_R was defined individually for each parton splitting using the CKKW prescription.

10.1.3 PYTHIA

PYTHIA delivers two different event generator models: one model based on a Q^2 -ordered parton-shower algorithm (referred to as the *old* model below), and one model based on a p_T -ordered parton-shower algorithm (referred to as the *new* model below). For the former model the configuration parameters were set according to tune QW, and for the latter according to tune S0 [134, 135].

10.1.4 HERWIG+JIMMY

The HERWIG sample was generated using v6.510 of the code together with JIMMY v4.31 with default configuration parameters.

10.1.5 MC@NLO+JIMMY

The MC@NLO sample was generated using release v3.3. HERWIG v6.510 provided the parton shower and hadronization models, whereas JIMMY v4.31 provided simulation of multiple parton interactions. All configuration parameters were set to their default values.

10.1.6 MCFM

MCFM v5.3 was used to produce fixed-order pQCD NLO (LO) predictions for the two (three) leading p_T^{jet} spectra for $Z/\gamma^*(\rightarrow e^+e^-)$ +jets. The internal cone-jet algorithm in MCFM was modified to correspond to the DØ Run II Mid-Point algorithm following Ref. [136]. The central predictions were generated with the factorization and renormalization scales set to $\mu_F = \mu_R = \sqrt{M_Z^2 + p_{T,Z}^2}$. The PDF uncertainties were estimated according to the Hessian method [124]. For the two leading jets, they vary from 5% at low p_T to 10% at high p_T , and for the third jet they are found to be (5–15)%.

The MCFM predictions were multiplied by correction factors accounting for multiple parton interactions (C_{MPI}) and hadronization (C_{Had}) before being compared to the measurements. The correction factors C_{MPI} and C_{Had} were estimated using inclusive $Z/\gamma^*(\rightarrow e^+e^-)$ event samples generated with Q^2 -ordered PYTHIA using Tune QW, p_T -ordered PYTHIA using Tune S0 [134, 135], HERWIG+JIMMY, ALPGEN+PYTHIA, as well as SHERPA. Each p_T^{jet} spectrum was evaluated for each generator both with the hadronization and multiple parton interaction model turned off (sample *a*), with only the multiple parton interaction model turned on (sample *b*), and with both models turned on (sample *c*). For each generator the ratio of the

Table 10.1: Correction factors for multiple parton interactions (C_{MPI}) and hadronization (C_{Had}) for $\frac{1}{\sigma_{Z/\gamma^*}} \times \frac{d\sigma}{dp_T(1^{\text{st}} \text{ jet})}$.

p_T bin [GeV]	$C_{\text{MPI}} \pm (\text{stat}) \pm (\text{sys})$	$C_{\text{Had}} \pm (\text{stat}) \pm (\text{sys})$
20–28	$1.08 \pm 0.00^{+0.07}_{-0.04}$	$0.89 \pm 0.00^{+0.04}_{-0.03}$
28–40	$1.04 \pm 0.00^{+0.02}_{-0.02}$	$0.90 \pm 0.00^{+0.03}_{-0.01}$
40–54	$1.02 \pm 0.00^{+0.01}_{-0.01}$	$0.90 \pm 0.00^{+0.02}_{-0.00}$
54–73	$1.02 \pm 0.01^{+0.00}_{-0.02}$	$0.92 \pm 0.01^{+0.01}_{-0.03}$
73–95	$1.01 \pm 0.01^{+0.03}_{-0.01}$	$0.93 \pm 0.01^{+0.01}_{-0.02}$
95–120	$1.02 \pm 0.02^{+0.00}_{-0.03}$	$0.91 \pm 0.02^{+0.03}_{-0.00}$
120–154	$1.04 \pm 0.03^{+0.00}_{-0.07}$	$0.92 \pm 0.02^{+0.05}_{-0.03}$
154–200	$1.03 \pm 0.05^{+0.02}_{-0.06}$	$0.91 \pm 0.04^{+0.04}_{-0.06}$
200–300	$1.01 \pm 0.09^{+0.04}_{-0.05}$	$0.92 \pm 0.08^{+0.05}_{-0.06}$

 Table 10.2: Correction factors for multiple parton interactions (C_{MPI}) and hadronization (C_{Had}) for $\frac{1}{\sigma_{Z/\gamma^*}} \times \frac{d\sigma}{dp_T(2^{\text{nd}} \text{ jet})}$.

p_T bin [GeV]	$C_{\text{MPI}} \pm (\text{stat}) \pm (\text{sys})$	$C_{\text{Had}} \pm (\text{stat}) \pm (\text{sys})$
20–28	$1.15 \pm 0.01^{+0.06}_{-0.10}$	$0.81 \pm 0.01^{+0.07}_{-0.00}$
28–40	$1.10 \pm 0.01^{+0.00}_{-0.07}$	$0.83 \pm 0.01^{+0.05}_{-0.00}$
40–54	$1.07 \pm 0.02^{+0.00}_{-0.06}$	$0.85 \pm 0.01^{+0.06}_{-0.00}$
54–73	$1.04 \pm 0.03^{+0.00}_{-0.07}$	$0.87 \pm 0.03^{+0.07}_{-0.01}$
73–200	$1.05 \pm 0.05^{+0.00}_{-0.08}$	$0.83 \pm 0.04^{+0.18}_{-0.00}$

observable in sample b to that in sample a specifies C_{MPI} . Correspondingly, the ratio of the observable in sample c to that in sample b specifies C_{Had} . The central values of C_{MPI} and C_{Had} given in Tables 10.1, 10.2 and 10.3 correspond to the predictions of PYTHIA using tune QW for the first, second and third p_T^{jet} spectrum, respectively. The maximal upwards and downwards differences between PYTHIA Tune QW and the other four models are quoted as systematic uncertainties. A reasonable level of agreement between the various event generators models is seen, indicating that the correction for hadronization and multiple parton interactions is well understood.

10.2 Comparing the Predictions with the p_T^{jet} Measurements

The measurements of the cross section for $Z/\gamma^*(\rightarrow e^+e^-)+N$ -jet production ($N = 1, 2, 3$), differential in the p_T of the N^{th} jet and normalized to the inclusive cross section, are shown in Fig. 10.1 together with the NLO (LO) predictions from MCFM for $N = 1, 2$ ($N = 3$). In this and later figures the uncertainty bars attached to the data points represent the statistical and

Table 10.3: Correction factors for multiple parton interactions (C_{MPI}) and hadronization (C_{Had}) for $\frac{1}{\sigma_{Z/\gamma^*}} \times \frac{d\sigma}{dp_T(3^{\text{rd}} \text{ jet})}$.

p_T bin [GeV]	$C_{\text{MPI}} \pm (\text{stat}) \pm (\text{sys})$	$C_{\text{Had}} \pm (\text{stat}) \pm (\text{sys})$
20–28	$1.15 \pm 0.02^{+0.00}_{-0.07}$	$0.76 \pm 0.01^{+0.08}_{-0.00}$
28–44	$1.10 \pm 0.03^{+0.05}_{-0.04}$	$0.81 \pm 0.03^{+0.05}_{-0.00}$
44–60	$1.11 \pm 0.10^{+0.04}_{-0.10}$	$0.74 \pm 0.07^{+0.19}_{-0.00}$

systematic uncertainties of the measurement combined in quadrature. The MCFM predictions are corrected to the particle level using the product ($C_{\text{MPI}} \times C_{\text{Had}}$) from Tables 10.1–10.3.

10.2.1 Fixed-order Calculations at LO and NLO from MCFM

To facilitate a quantitative comparison between data and the theory predictions, the ratios of data and each prediction to a reference prediction is studied. The reference prediction is taken to be that of NLO (LO) MCFM for the two leading (third) jets, with $\mu_F = \mu_R = \sqrt{M_Z^2 + p_{T,Z}^2}$. In Fig. 10.2 the resulting ratios are shown for data and the NLO MCFM prediction for the nominal scale choice as well as for both μ_F and μ_R varied up and down by a factor of two (keeping $\mu_F = \mu_R$). Reasonable agreement is seen between the NLO MCFM predictions and the measurements, both for the leading and the second leading jets. The variation in scales leads to a variation in the predicted differential cross section of ~ 1.1 both for the leading and second leading jet. The central LO predictions are about 20–30% below the measurements, but the shapes of the distributions are well described (see Fig. 10.3). The scale uncertainty is about 1.2 / 1.4 / 1.6 for the first / second / third jets, indicating a significantly lower predictive power for the LO calculation than for the NLO calculations.

10.2.2 SHERPA and ALPGEN+PYTHIA

In contrast to parton-level predictions from fixed-order calculations, the particle-level predictions of event-generator models can be used as input for a detailed simulation of the experimental detector setup. A good understanding of the detector performance is essential for an accurate interpretation of experimental data samples. It is therefore of great importance to both measurements and searches for new particles to have access to reliable particle-level event-generator models. As described in Chapter 3, the SHERPA and ALPGEN+PYTHIA event generator models combine tree-level $2 \rightarrow n$ matrix-element predictions with a parton-shower algorithm, aiming to deliver more accurate predictions for events with one or more QCD jets than achievable using a parton-shower algorithm alone. Both event generators are found to give a reasonable description of the shapes of the measured p_T^{jet} spectra in data, as seen in Figs. 10.4 and 10.5. However, neither generator describes the normalizations of the distributions correctly, with SHERPA (ALPGEN) predicting more (less) jets than observed in data. Given that SHERPA and ALPGEN use LO tree-level matrix-element calculations to describe hard, well-separated jets, one would naively expect these generators to reproduce the p_T^{jet} spectra predicted by LO MCFM. This is indeed the case for ALPGEN+PYTHIA, but

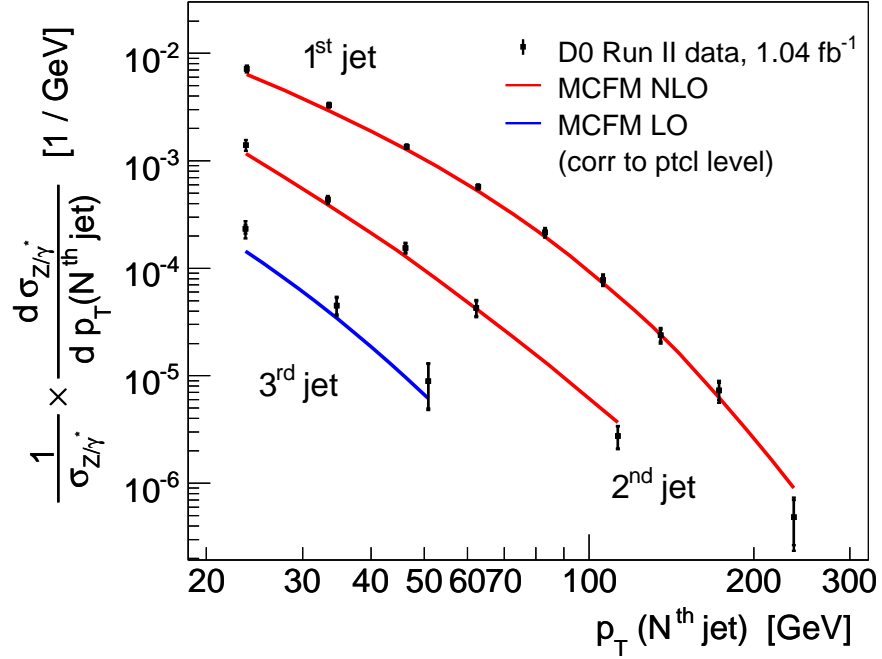


Figure 10.1: Comparison between the measured p_T^{jet} spectra and the predictions of MCFM for $p_T(N^{\text{th}} \text{ jet})$ for $Z/\gamma^*(\rightarrow e^+e^-)+N\text{-jet}$ production, $N = 1, 2, 3$. The MCFM predictions are corrected for hadronization and the underlying event.

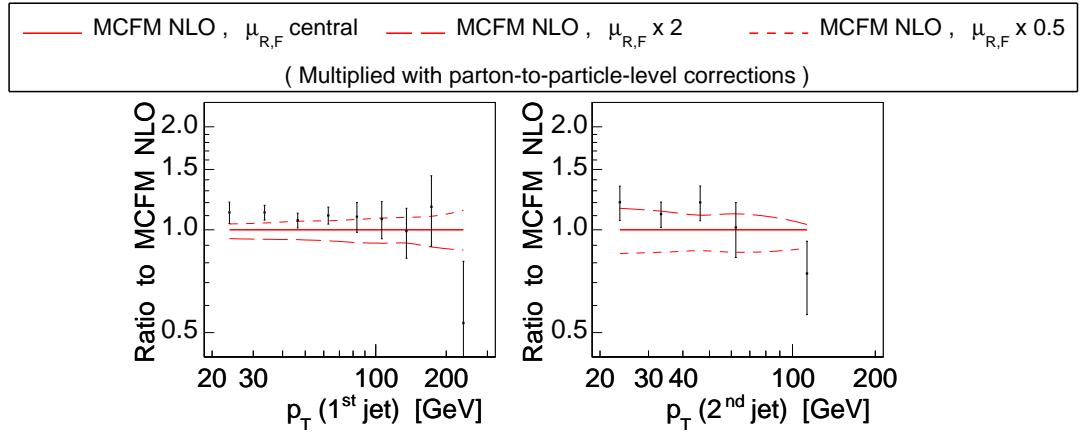


Figure 10.2: Ratios of the two leading p_T^{jet} spectra measured in data and predicted by NLO MCFM to the reference prediction (NLO MCFM, central scales). Three ratios are shown for MCFM, corresponding to the nominal scale choice $\mu_F = \mu_R = \sqrt{M_Z^2 + p_{T,Z}^2}$, as well as to this choice multiplied by 1/2 and 2.

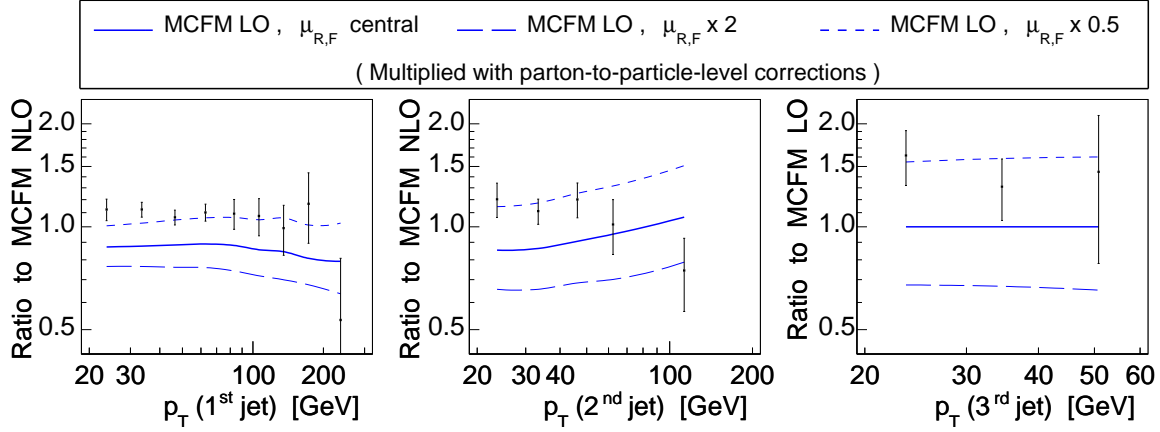


Figure 10.3: Ratios of the p_T^{jet} spectra measured in data and of those predicted by LO MCFM to the reference predictions. For LO MCFM three ratios are shown, corresponding to the nominal scale choice $\mu_F = \mu_R = \sqrt{M_Z^2 + p_{T,Z}^2}$, as well as to this choice multiplied by 1/2 and 2.

SHERPA predicts significantly more jets than LO MCFM. As expected, the scale uncertainties of both event generators are comparable to those of LO MCFM. The ALPGEN code also offers the possibility of using HERWIG+JIMMY instead of PYTHIA for providing the parton shower, hadronization model and description of the underlying event, but both choices are found to produce similar predictions (see Fig. 10.5).

10.2.3 MC@NLO

An alternative method for combining matrix-element calculations with a parton-shower algorithm is implemented in the MC@NLO generator. This generator combines a NLO matrix-element calculation for the core process, in this case $q\bar{q} \rightarrow Z/\gamma^* \rightarrow e^+e^-$, with a parton-shower algorithm. In other words, the properties of the leading jet is generated with LO matrix-element prediction, whereas sub-leading jets are described by the parton-shower algorithm. As seen in Fig. 10.6 the MC@NLO predictions deviate significantly from data for all three p_T^{jet} spectra. For the leading jet MC@NLO predicts a factor of 0.8 to 0.5 less jets than observed in data, depending on p_T^{jet} . For the sub-leading jets very large disagreements are seen, both in shape and normalization. The origins of these discrepancies are discussed below.

10.2.4 PYTHIA and HERWIG

Event generators relying on a parton-shower algorithm include, most notably, PYTHIA and HERWIG. While originally delivering pure parton-shower predictions, current version of both generators by default reweight the leading initial-state parton-shower emission for selected processes according to LO matrix-element predictions². The predictions of the Q^2 -ordered PYTHIA model are compared with data in Fig. 10.7, both for the central scale choices of $\mu_F = M_Z$, $\mu_R = p_T$, where p_T denotes the relative transverse momentum of the $1 \rightarrow 2$

²For PYTHIA, such reweighting is offered for the s -channel production of a single gauge or Higgs boson [50].

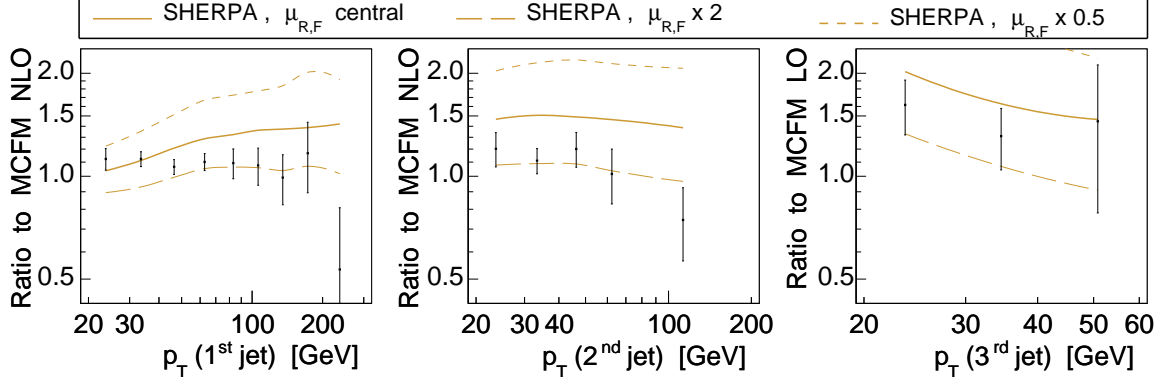


Figure 10.4: Ratios of the measurements and of the predictions of SHERPA to the reference predictions from MCFM. For SHERPA three ratios are shown, corresponding to the nominal scale choice, as well as to the nominal choice of factorization and renormalization scales, as well as to the nominal scale choice multiplied by 1/2 and 2.

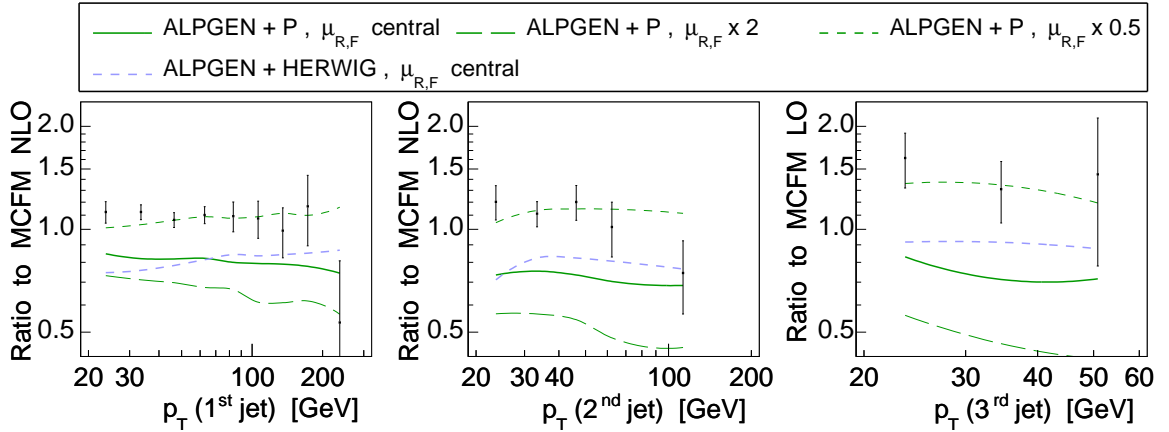


Figure 10.5: Ratios of the p_T^{jet} measurements and of the predictions of ALPGEN+PYTHIA to the reference predictions from MCFM. For ALPGEN+PYTHIA three ratios are shown, corresponding to the nominal choice of factorization and renormalization scales, as well as to the nominal scale choice multiplied by 1/2 and 2.

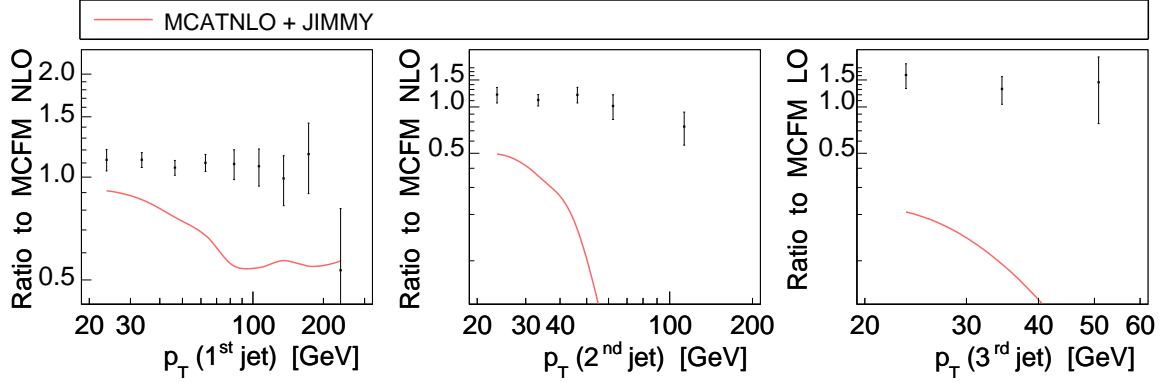


Figure 10.6: Ratios of the p_T^{jet} measurements and of the predictions of MC@NLO to the reference predictions from MCFM.

splitting generated by the parton shower, and for both scales varied by a factor of two up and down. For the leading jet the PYTHIA prediction is comparable to that of LO pQCD, modulo a minor disagreement in shape. This agrees with the fact that the leading emission has been reweighted to reproduce a LO Z/γ^* +parton calculation. For the sub-leading jets significant disagreements with both data and the NLO MCFM predictions are seen, both for normalization and shape. This is in agreement with what one would expect from the limited validity of the soft/collinear approximation for events with several well-separated jets with large values of p_T . The newer, p_T -ordered PYTHIA model is found to offer an improved description of the leading jet (see Fig. 10.8). Since the leading p_T^{jet} spectrum is determined by the the matrix-element reweighting algorithm rather than by the parton-shower algorithm, it is reasonable to assume that the improved description of the leading p_T^{jet} spectrum is due to a modification of the former algorithm rather than due to the change of parton-shower evolution variable. For the sub-leading p_T^{jet} spectra no improvement over the Q^2 -ordered model is seen.

The HERWIG event generator model, using an angular-ordered parton-shower algorithm, offers predictions which are in general similar to those of PYTHIA (see Fig. 10.9). One notable feature is a change of slope in the p_T spectrum of the leading jet at approximately $p_T = 50$ GeV. Below this value HERWIG agrees well with data, whereas above this value the prediction is similar to that of the Q^2 PYTHIA model. This feature can be traced back to the technical details of the matrix-element reweighting used in HERWIG [137]: different reweighting schemes are used for small and large values of p_T^{jet} , and for $Z/\gamma^* \rightarrow e^+e^-$ production the transition region is at $p_T^{\text{jet}} \approx 50$ GeV. This illustrates that the $p_T(\text{leading jet})$ spectrum for this class of generators is determined by the details of the matrix-element reweighting, not by the parton-shower algorithm.

10.2.5 The Relationship between MC@NLO, SHERPA, PYTHIA and HERWIG

It can be argued that PYTHIA and HERWIG should reproduce MC@NLO for $p_T(\text{leading jet})$ since all three are formally based on LO matrix-element predictions for this observable. The same is true for SHERPA (or ALPGEN+PYTHIA) in case only up to one QCD parton is included in the matrix-element calculation (labelled $N_{\text{jet}} = 1$ below). In Fig. 10.10 the predictions

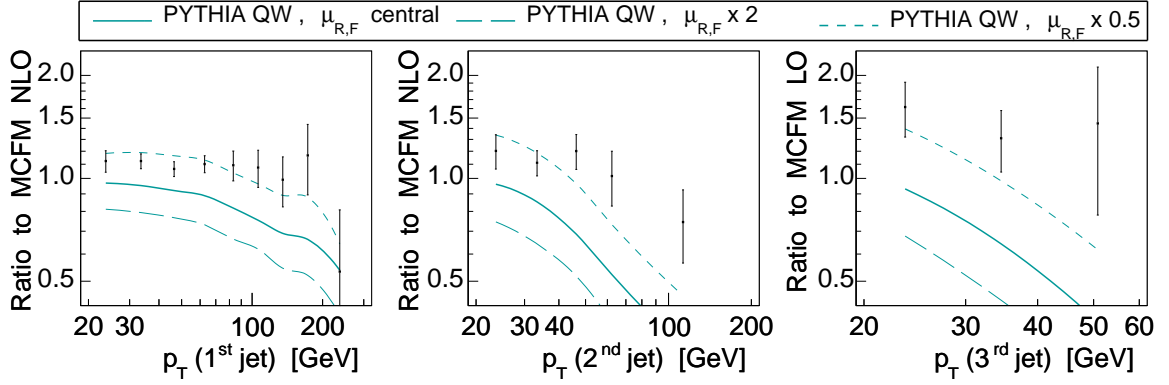


Figure 10.7: Ratios of the p_T^{jet} measurements and of the predictions of Q^2 -ordered PYTHIA using tune QW to the reference prediction from MCFM. Three ratios are shown for PYTHIA, corresponding to the nominal choice of factorization and renormalization scales and to the nominal scale choice multiplied by 1/2 and 2.

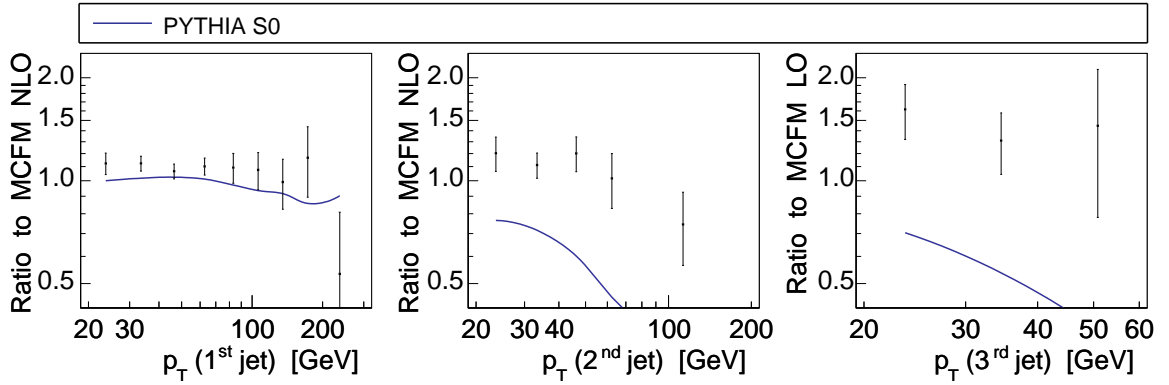


Figure 10.8: Ratios of the p_T^{jet} measurements and of the predictions of p_T -ordered PYTHIA using tune S0 to the reference prediction from MCFM.

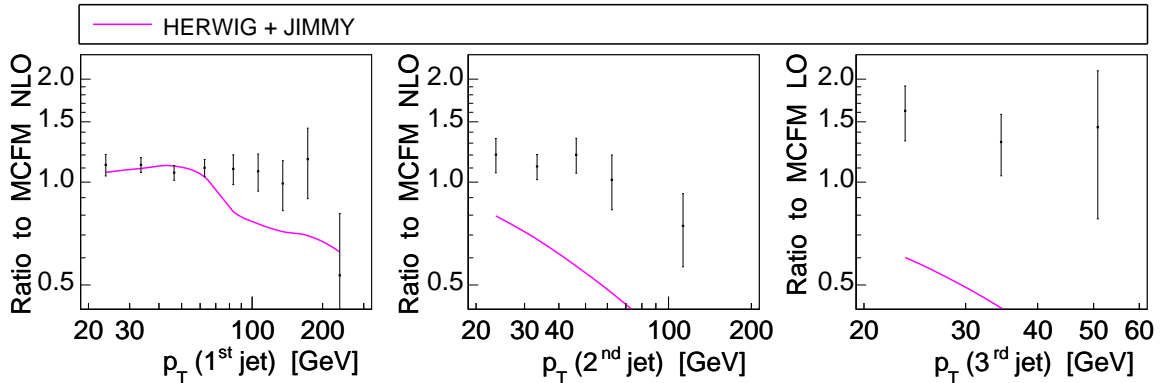


Figure 10.9: Ratios of the p_T^{jet} measurements and of the predictions of HERWIG+JIMMY to the reference predictions from MCFM.

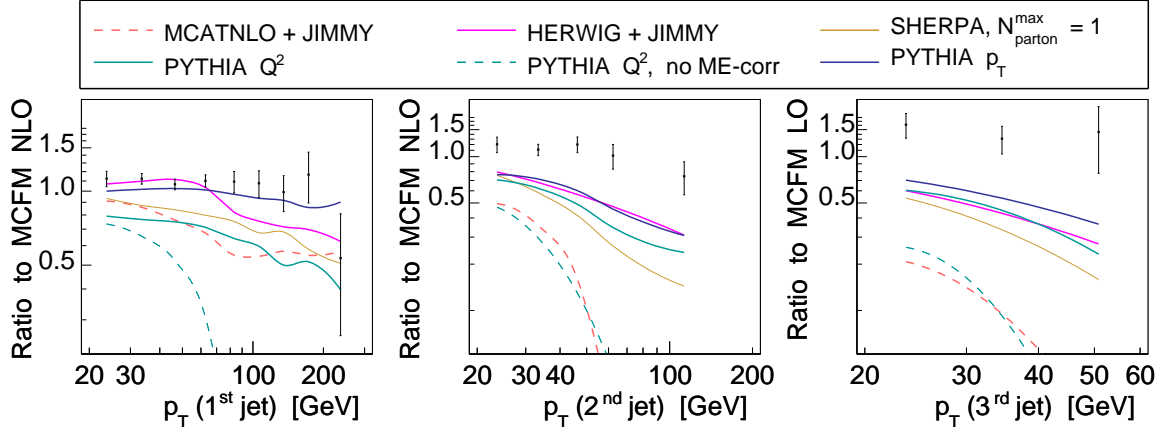


Figure 10.10: Ratios of the p_T^{jet} measurements and of the predictions of several different event-generator models (see legend) to the reference predictions from MCFM. An interpretation of the relative performance of the different models is given in the main text.

of HERWIG and the two PYTHIA models are compared with both MC@NLO and SHERPA with $N_{\text{jet}} = 1$. To highlight the impact of the matrix-element reweighting in PYTHIA, predictions are also shown with this reweighting turned off, with the starting-scale of the shower set to M_Z^2 . Several observations can be made: first, one class of event generator models which are all in reasonable agreement with each other, for all three p_T^{jet} spectra, consists of the matrix-element reweighted HERWIG and PYTHIA models as well as SHERPA with $N_{\text{jet}} = 1$. Second, disabling the matrix-element reweighting in PYTHIA leads to a dramatic reduction in the amount of events with one or more hard jets. Third, MC@NLO is found to be in reasonable agreement with the former class of event generator model for the leading jet, but accurately reproduces PYTHIA without reweighting for the two sub-leading jets. This behaviour is assumed to reflect that both of these models uses M_Z^2 as starting scale for the parton-shower algorithm, thereby placing a strict upper limit on the p_T values of the jets generated by the parton-shower algorithm. In contrast, the PYTHIA and HERWIG models, with matrix-element reweighting turned on, uses the full center-of-mass energy as starting scale. The leading emission is modified to reproduce LO matrix-element predictions, but sub-leading emissions are not modified, meaning that a larger p_T^{jet} values can be generated by these models for the sub-leading jets than is the case for MC@NLO. The SHERPA model sets the parton-shower starting scale for each parton generated according to the matrix-element calculation to the scale where the parton was emitted, as defined through the k_T clustering part of the CKKW algorithm. For parton multiplicities beyond the maximal number of partons included in the matrix-element calculations (N_{jet}), the veto of hard, well-separated parton-shower emissions is disabled. For $N_{\text{jet}} = 1$ this is seen to give predictions which are similar to those of PYTHIA and HERWIG for all three p_T^{jet} spectra.

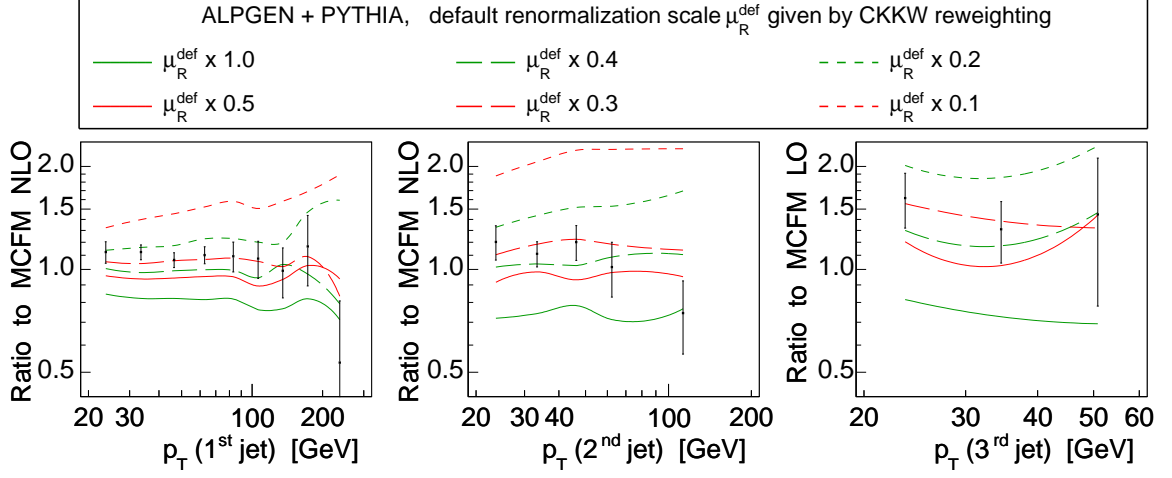


Figure 10.11: Ratios of the measurements and the predictions from ALPGEN+PYTHIA to the reference predictions from MCFM. The various ALPGEN+PYTHIA predictions correspond to the nominal choice of renormalization scale, as well as to the nominal choice multiplied with a factor of 0.5, 0.4, 0.3, 0.2 or 0.1.

10.2.6 Adjusting the parameters of ALPGEN+PYTHIA and SHERPA to describe the p_T^{jet} Measurements

The systematic uncertainties of the SHERPA and ALPGEN+PYTHIA predictions arising from variations of the unphysical parameters of the models were evaluated in Ref. [41]. Given the p_T^{jet} spectra measured in data, it is useful to study to which extent the parameters of the models can be modified in order to improve the agreement with data. This process is often referred to as *tuning* the parameters of the model. One adjustable parameter of all perturbative calculations is the renormalization scale. Due to the running of α_s , a lower value of μ_R implies a larger value of α_s , and consequently higher jet cross sections. The predictions for ALPGEN+PYTHIA and SHERPA for different choices of μ_R are given in Figs. 10.11 and 10.12, respectively. For ALPGEN+PYTHIA a good, simultaneous agreement with all three measured p_T^{jet} spectra is achieved by scaling the nominal value of μ_R by a factor of ~ 0.3 . For SHERPA, a comparable level of agreement is reached by multiplying the nominal μ_R by a factor of approximately 2, but some residual disagreements are seen in the shape of the p_T spectrum of leading jet.

The CKKW and MLM algorithms both contain two parameters which separates the parton which are generated according to matrix-element calculations from those which are generated by the parton-shower algorithm (see Sects. 3.3.2 and 3.3.2). In the case of CKKW, as implemented in SHERPA, this boundary is defined using the k_T jet algorithm which contains two parameters p_T^{cut} and D (defined in Sect. 3.7.2). These parameters offer an alternative to the renormalization scale when trying to tune the event generator models to agree with data. The variation in the p_T^{jet} spectra arising from reducing D from the default value of 1.0 to 0.75 and 0.5 is found to be negligible (see Fig. 10.13). Varying the p_T^{cut} parameter from the nominal value of 15 GeV to 10 GeV and 20 GeV leads to a change in normalization for the two leading p_T^{jet} spectra on the 10% level, as seen in Fig. 10.14. Choosing a p_T^{cut} value which is larger than 20 GeV results in significant changes in both normalization and shape. Notice

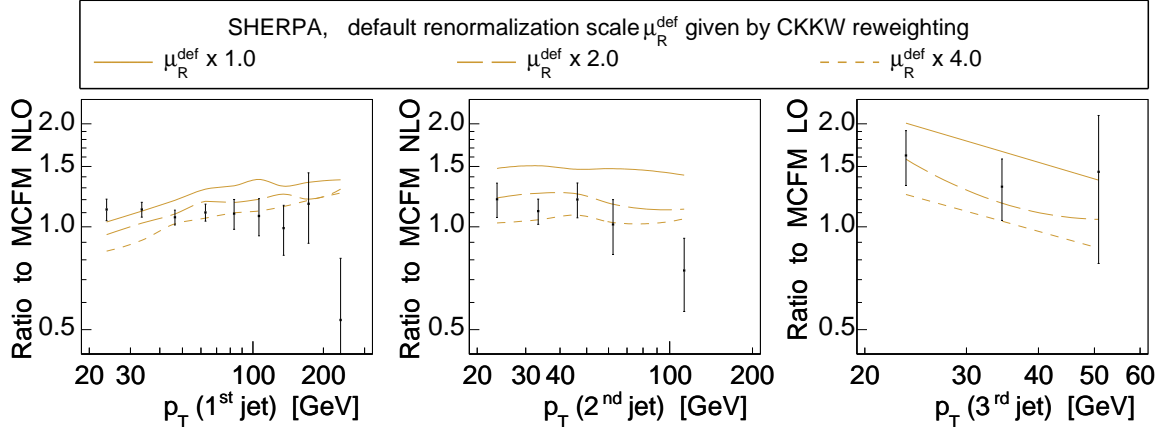


Figure 10.12: Ratios of the measurements and of the predictions from SHERPA to the reference predictions from MCFM. The various SHERPA predictions correspond to the nominal choice of renormalization scale, as well as to the nominal choice multiplied with a factor of 2 and 4.

that this is to be expected since a large p_T^{cut} value implies that well-separated, high- p_T partons are generated by the parton-shower algorithm, and for such partons the soft/collinear approximation of the parton shower is not a good choice. The predictions corresponding to large values of p_T^{cut} are therefore not meant to reflect the underlying uncertainty of the model and should be interpreted merely as an illustration of the behaviour of the CKKW algorithm in the limit of large p_T^{cut} . For $p_T^{\text{cut}} = 100$ GeV the predicted spectrum is found to be similar to that of PYTHIA without matrix-element correction applied and the parton-shower starting scale set to M_Z^2 . Starting from about $p_T = 100$ GeV the partons are generated according to the matrix-element calculation, and consequently there is a rapid increase in the ratio to MCFM. This illustrates that the value of p_T^{cut} should not exceed the range where the soft/collinear parton-shower approximation delivers reasonable predictions. Whereas the value $p_T^{\text{cut}} = 60$ GeV gives an improved level of agreement with the NLO MCFM predictions for the leading p_T^{jet} spectrum, large discrepancies are seen for the sub-leading jets. This indicates that p_T^{cut} is a less suited parameter for tuning the event generator to reproduce data or higher-order calculations than the renormalization scale parameter.

10.3 Comparing the Predictions with the $\Delta\phi(\text{di-electron, jet})$ Measurements

While fixed-order calculations were found to be well suited for predicting p_T^{jet} spectra, the same is not true for the $\Delta\phi(\text{di-electron, jet})$ observable. For $\Delta\phi \approx \pi$ the contributions of logarithmically enhanced higher-order terms corresponding to multiple emission of soft gluons become important, and an analytical resummation or parton-shower algorithm is necessary to describe this region of phase space. The shape of the $\Delta\phi$ spectrum measured in data is represented graphically in Fig. 10.15, together with the prediction from SHERPA, and good agreement is seen. Ratios of data and various event generator predictions with respect to the SHERPA predictions are presented in Fig. 10.16. For SHERPA and ALPGEN+PYTHIA the

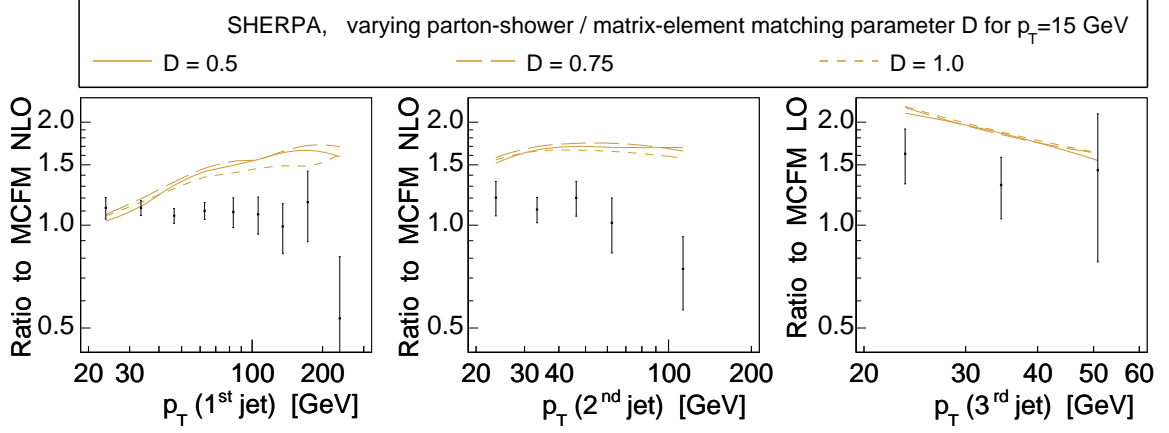


Figure 10.13: Ratios of the measurements and of the predictions from SHERPA to the reference predictions from MCFM. The various SHERPA predictions were generated using the nominal choice of the matching parameter $D = 1.0$, as well as $D = 0.75$ and $D = 0.5$.

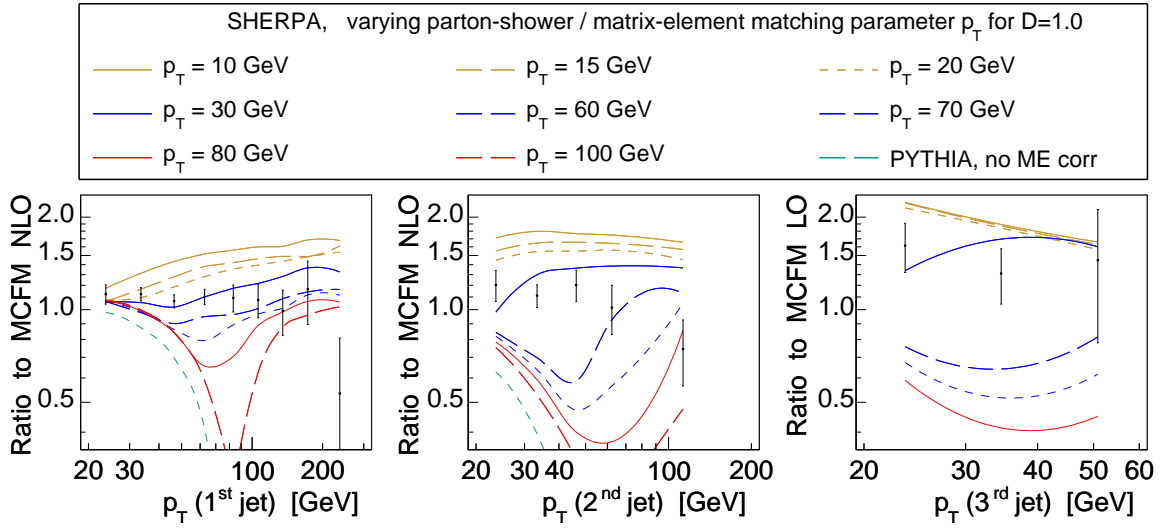


Figure 10.14: Ratios of the measurements and of the predictions from SHERPA to the reference predictions from MCFM. The various SHERPA predictions were generated using the nominal choice of the matching parameter $p_T^{\text{cut}} = 15$ GeV, as well as the values 10 GeV, 20 GeV, 30 GeV, 60 GeV, 70 GeV, 80 GeV and 100 GeV. The values above ~ 30 GeV are outside of the range of reasonable values for this parameter and are included as an illustration of the CKKW algorithm only (see main text).

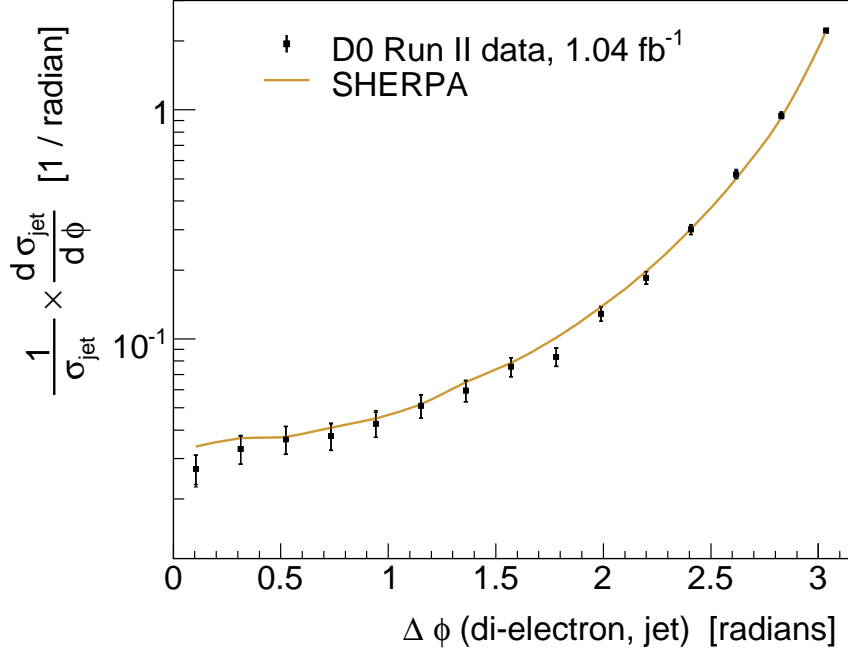


Figure 10.15: The distribution of $\Delta\phi(\text{di-electron, jet})$ (normalized to unity) as measured in data and as predicted by SHERPA.

distributions are also shown for the default values of μ_F and μ_R shifted up and down by a factor of two. It is seen that in particular the region close to π is sensitive to the scale choice. In contrast to the p_T^{jet} spectra, the nominal scale choice gives the best agreement between SHERPA and data. The spectrum predicted by ALPGEN+PYTHIA is more strongly peaked at π than the one measured in data. The region close to π is described by events where only one parton was included in the matrix-element calculation, as seen in Fig. 10.17. At the matrix-element level such events all have $\delta\phi = \pi$, but the parton-shower algorithm, as well as to a smaller extend the hadronization and intrinsic p_T models, lead to values smaller than π .

For the region $\Delta\phi < 2.5$, which is populated mostly by events with two parton included in the matrix-element calculation, the shape of the ALPGEN+PYTHIA prediction agrees well with data. Consequently, the source of the discrepancies between data and ALPGEN+PYTHIA is most likely to be found in the parton-shower algorithm rather than in the matrix-element part of the event generator model. Varying μ_F and μ_R down by a factor of two results in an improved level of agreement with data. The two PYTHIA models, HERWIG and MC@NLO all show disagreements in shape with respect to data over the whole range $[0, \pi]$.

10.4 The Sensitivity of $\Delta\phi(\text{jet, jet})$ to Jets from Multiple Parton Interactions

In Chapter 7 a detector-level comparison of SHERPA and Q^2 -ordered PYTHIA with data for the distribution of the difference in azimuthal angle between the two leading jets in

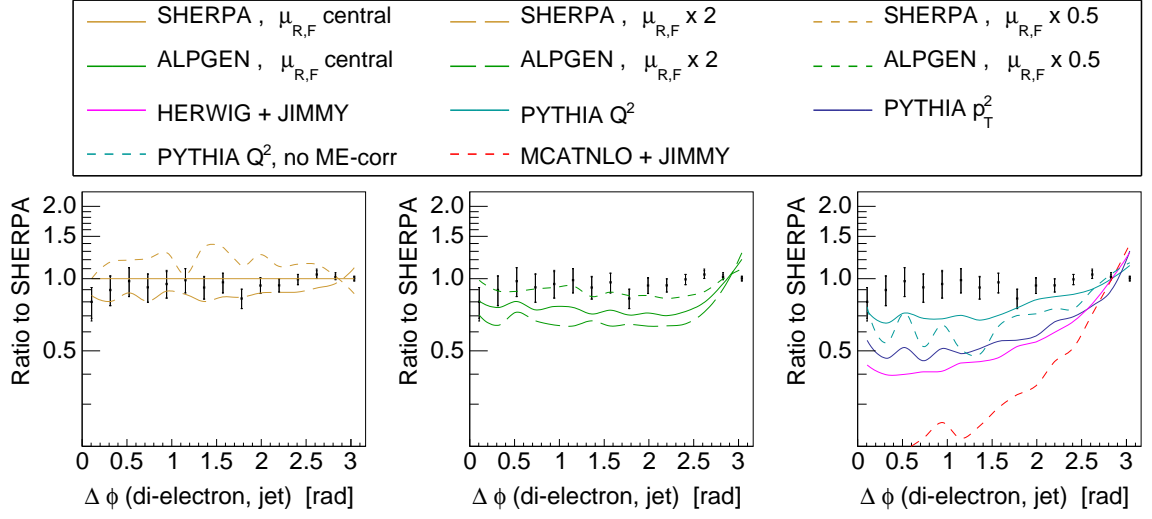


Figure 10.16: Ratios of the $\Delta\phi(\text{di-electron},\text{jet})$ measurement and of the predictions of several different event-generator models (see legend) to the reference prediction (SHERPA, central scales). An interpretation of the relative performance of the different models is given in the main text.

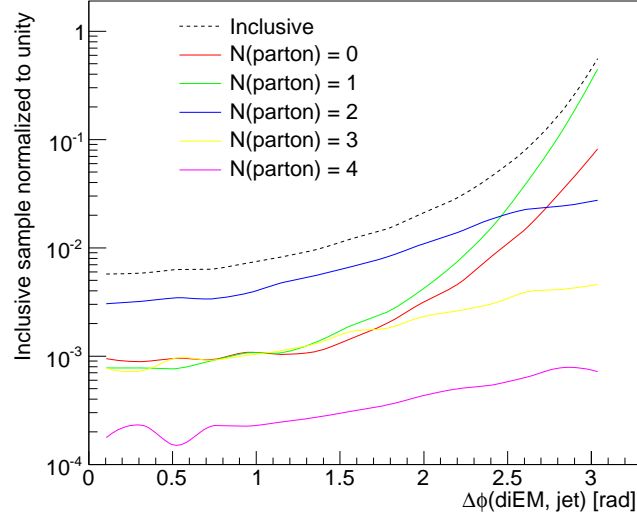


Figure 10.17: The $\Delta\phi(\text{di-electron},\text{jet})$ spectrum (normalized to unity) as predicted by ALPGEN+PYTHIA, shown once for the inclusive sample, and for all events where N partons were included in the matrix-element calculation ($N = 0, 1, 2, 3, 4$).

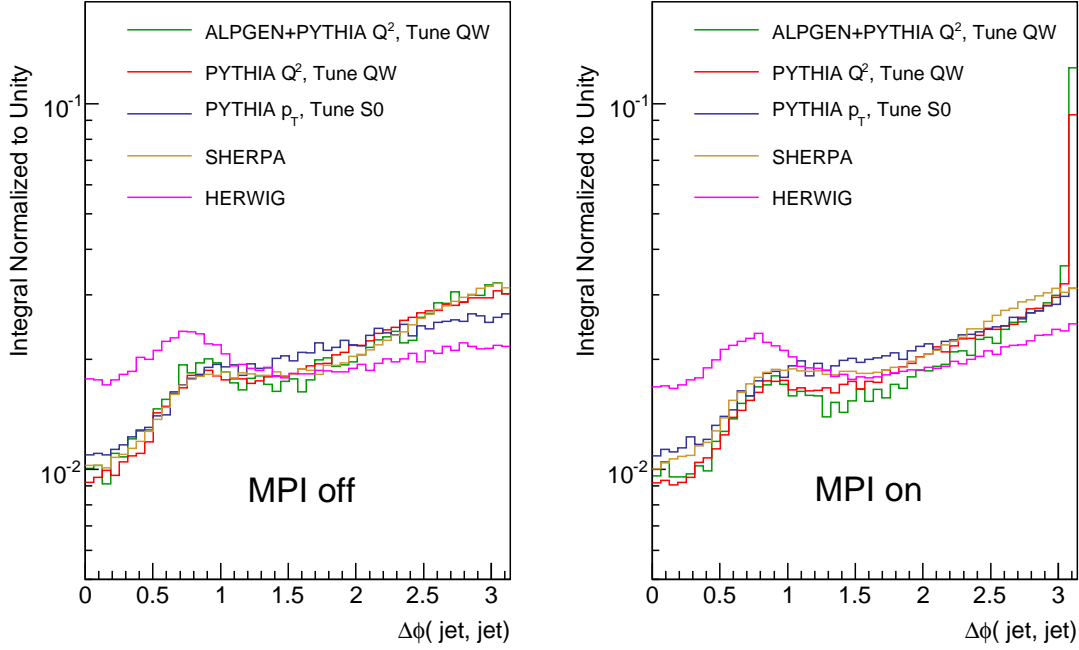


Figure 10.18: The $\Delta\phi(\text{jet},\text{jet})$ spectrum (normalized to unity) as predicted by several different event generator codes without (left) and with (right) the simulation of multiple parton interactions. A prominent excess of events at $\Delta\phi = \pi$ is seen to be generated by the multiple parton interaction model for Q^2 -ordered PYTHIA. This peak was not seen in the $\Delta\phi(\text{jet},\text{jet})$ spectrum reconstructed in data (see Fig. 7.12).

$Z/\gamma^* \rightarrow e^+e^- + \text{jets}$ events was presented (see Fig. 7.12). In the PYTHIA spectrum a significant excess of events with a $\Delta\phi(\text{jet},\text{jet}) \approx \pi$ was seen, a feature which was not present in the spectrum reconstructed in data, nor in the predictions from SHERPA. In Fig. 10.18 the $\Delta\phi(\text{jet},\text{jet})$ distribution as predicted by the following event generators is shown: Q^2 -ordered PYTHIA, p_T -ordered PYTHIA, ALPGEN+PYTHIA (Q^2 -ordered), SHERPA and HERWIG. For each generator the spectrum was generated with the simulation of multiple parton interactions both activated and deactivated. The excess of events with $\Delta\phi(\text{jet},\text{jet}) \approx \pi$ is only seen in the two samples generated using the multiple parton interaction model of Q^2 -ordered PYTHIA. Based on this, and on the detector-level comparisons between PYTHIA and data, it can be concluded that the multiple parton interaction model implemented in Q^2 -ordered PYTHIA is in disagreement with experimental data. In contrast to the multiple parton interactions model used in SHERPA, HERWIG and in the p_T -ordered PYTHIA model, the multiple parton interaction model used in Q^2 -ordered PYTHIA generates multiple parton interactions according to $2 \rightarrow 2$ QCD processes without the inclusion of higher-order corrections via a parton-shower algorithm [50]. Consequently, all parton pairs generated by this multiple parton interaction model are generated with $\Delta\phi(\text{jet},\text{jet}) = \pi$, thereby explaining the large excess of events seen in the bin corresponding to this value.

Chapter 11

From the Tevatron to the LHC

At the time of writing the Tevatron accelerator still delivers the particle collisions with the highest center-of-mass energy in the world. However, in the near future this distinction is expected to belong to the Large Hadron Collider (LHC) located at the boundary between France and Switzerland, close to Geneva. In the present Chapter it is highlighted how the differential p_T^{jet} spectra for the associated production of Z/γ^* and jets at the LHC differs from those at the Tevatron. In addition, it is investigated how the predictions of fixed-order NLO calculation as well as from SHERPA and ALPGEN+PYTHIA extrapolate from Tevatron $p\bar{p}$ collisions at 1.96 TeV to LHC pp collisions at 10 TeV.

11.1 p_T^{jet} spectra

The fixed-order NLO prediction for the jet-inclusive $Z/\gamma^* \rightarrow e^+e^-$ cross section, as predicted by MCFM using the CTEQ 6.1M PDFs, is 1300_{-26}^{+39} pb, where the quoted uncertainties correspond to a variation of the factorization and renormalization scales by a factor of two around the nominal value $\sqrt{M_Z^2 + p_{T,Z}^2}$. This is an increase by a factor of ~ 5 compared with the corresponding prediction for the Tevatron which is $238.6_{-2.6}^{+3.7}$ pb. The increase in cross section is even larger for events which contain one or more hard jets. In Fig. 11.1 the predicted $\frac{1}{\sigma_{Z/\gamma^*}} \times \frac{d\sigma}{dp_T}$ distributions for the three leading jets in Tevatron and LHC collisions are compared. The LHC (Tevatron) predictions are normalized using the inclusive $Z/\gamma^* \rightarrow e^+e^-$ cross section σ_{Z/γ^*} predicted for the LHC (Tevatron). It is seen that the fraction of inclusive $Z/\gamma^* \rightarrow e^+e^-$ events which contains at least one (two / three) jets with p_T larger than ~ 200 GeV (~ 100 GeV / ~ 40 GeV) is expected to be a factor of ten larger at the LHC than at the Tevatron. This large increase in the jet cross sections means that an accurate modeling of Z/γ^* +jets production will be even more important at the LHC than at the Tevatron. Given the success of NLO calculations in reliably predicting both the shape and normalization of the two leading p_T^{jet} spectra at the Tevatron, the recent appearance of NLO calculations for W +3-jet production [21, 22, 23] is an important milestone. The extension of these calculations to Z +3-jet production is expected to be straight forward, and it is foreseen [138] that such a calculation will soon be compared with the p_T (third jet) measurement presented in this thesis.

The p_T^{jet} spectra shown in Fig. 11.1 are the parton-level predictions from MCFM, with no correction for hadronization or the underlying event applied. Whereas hadronization is assumed to be well understood in terms of extrapolation from the Tevatron to the LHC, this is not the case for multiple parton interactions (see Chapter. 3). In Fig. 11.2, the SHERPA predictions for the normalized p_T^{jet} spectra for LHC (Tevatron) are presented as a ratio to the MCFM prediction for LHC (Tevatron) calculated at the NLO (LO) for the two leading jets (third jet). For both colliders samples were generated with the multiple

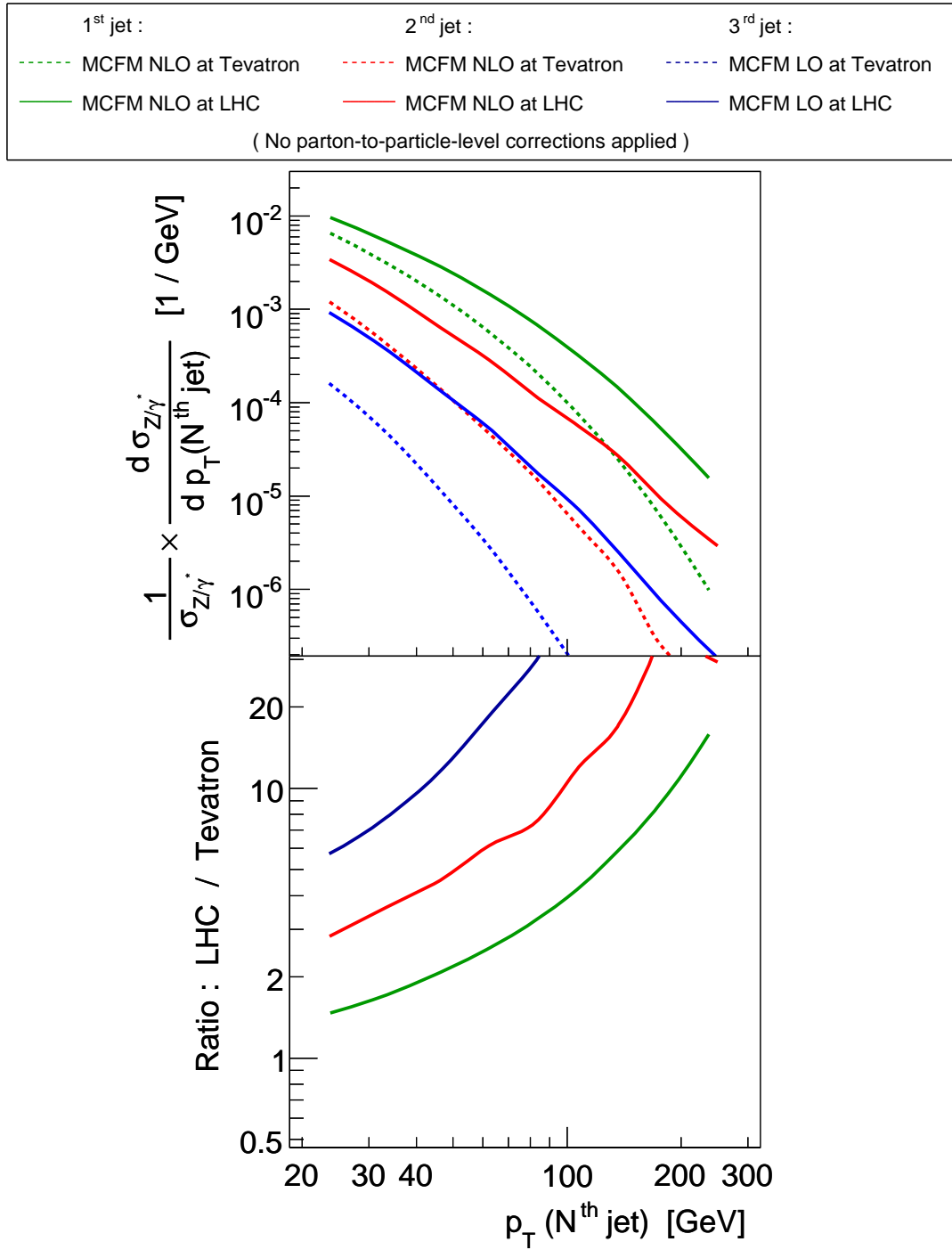


Figure 11.1: The p_T^{jet} spectra in $Z/\gamma^*(\rightarrow e^+e^-)+\text{jets}$ events at LHC compared with those at Tevatron, as predicted by fixed-order calculations from MCFM. The differential cross sections are shown relative to the predicted jet-inclusive $Z/\gamma^* \rightarrow e^+e^-$ cross sections. No corrections for hadronization of the underlying event was applied.

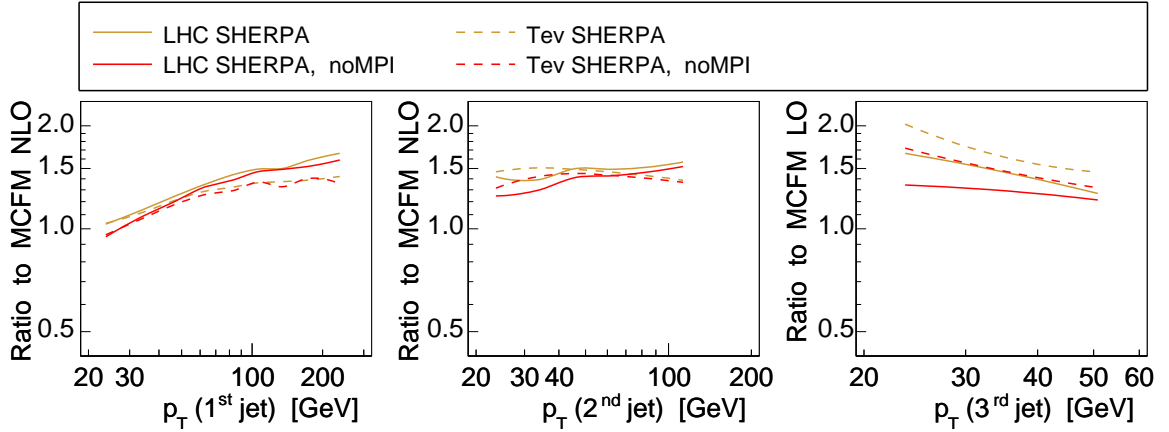


Figure 11.2: The ratio of the p_T^{jet} spectra predicted by SHERPA, with and without the simulation of multiple parton interactions, to those predicted by MCFM. The ratios are shown for both the Tevatron and the LHC.

parton interaction models turned on and off. The impact of activating the SHERPA multiple parton interaction model is seen to be similar both for the Tevatron and LHC predictions. Also, good agreement for the ratios to MCFM is seen between the SHERPA predictions for the two colliders, showing that SHERPA and the fixed-order MCFM calculations extrapolate similarly from Tevatron collisions to LHC collisions. The corresponding predictions from ALPGEN+PYTHIA using tune QW are given in Fig. 11.3. In contrast to SHERPA, this generator predicts a significant contribution from jets arising from multiple parton interactions at the LHC for p_T^{jet} below about 50 GeV. In absence of multiple parton interactions, the predicted ratios to the MCFM predictions are found to agree well between LHC and Tevatron. The large difference in the amount of multiple parton interactions generated by SHERPA and Q^2 -ordered PYTHIA was studied in Ref. [120]. By adjusting the parameters of the models in order to reproduced studies of the density of charged particles in dijet events [54] it was found that the SHERPA (Q^2 -ordered PYTHIA using tune A) model needs on average 1.08 (6.35) multiple parton interactions per event to describe data. This difference was interpreted as reflecting the lack of hard, multi-jet events in the perturbative part of the PYTHIA calculation due to the soft/collinear approximation in the parton shower. In contrast, SHERPA predicts a higher density of charged particles stemming from the perturbative calculation, thereby requiring a smaller number of multiple parton interactions to describe the data. Based on this, the parameters of the underlying event model in PYTHIA should be rederived using the full ALPGEN+PYTHIA model, but to the knowledge of the author such a set of parameter values is not currently available.

In the previous Chapter, the choice of renormalization scale for SHERPA and ALPGEN+PYTHIA were varied in order to improve the agreement with the p_T^{jet} spectra measured in data. Once corresponding measurements are available from the LHC experiments a similar optimization of parameters can be performed. For the Tevatron fixed-order NLO predictions were found to offer both the best level of agreement with data, and the highest predictive power. The prediction of NLO calculations for the LHC can therefore be assumed to deliver the best currently available estimation of the p_T^{jet} spectra for $Z/\gamma^*(\rightarrow e^+e^-)+\text{jet}$ production at LHC, and these predictions can be used to optimize the choice of renormalization scale used for SHERPA

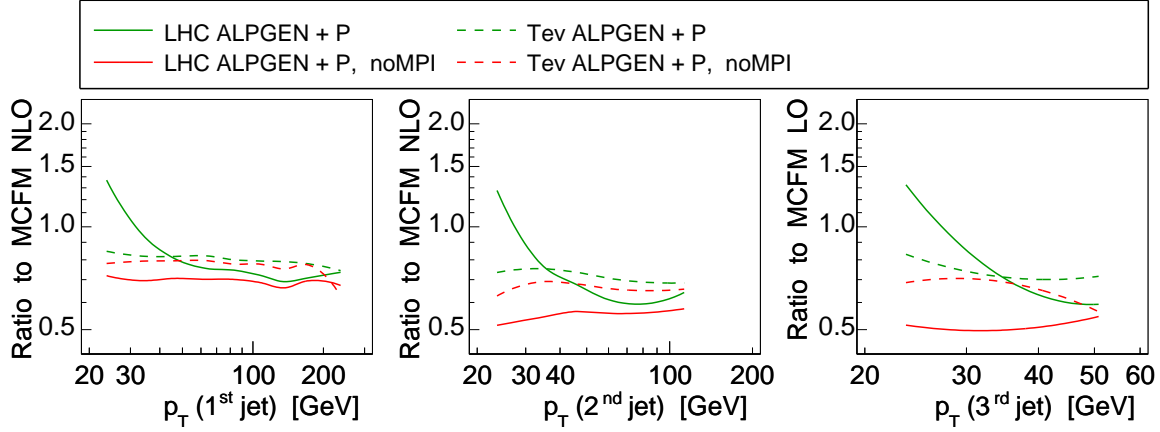


Figure 11.3: The ratio of the p_T^{jet} spectra predicted by ALPGEN+PYTHIA using tune QW (labelled ALPGEN+P), with and without the simulation of multiple parton interactions, to those predicted by MCFM. The ratios are shown for both the Tevatron and the LHC. At the LHC, a sizable contribution arising from multiple parton interactions is predicted for $p_T^{\text{jet}} < 50$ GeV.

and ALPGEN+PYTHIA. For the following studies the hadronization and multiple parton interaction models in SHERPA and ALPGEN+PYTHIA were deactivated. In Fig. 11.4 the ratio of the p_T^{jet} spectra predicted for LHC by SHERPA to those from MCFM are given for the nominal choice of μ_R as well as for upwards variations by factors of two and four. The latter choice is found to give a ratio close to unity for the second and third jets, but a residual difference in slope is seen for the leading jet. An upward variation of μ_R by a factor of four was also found to give good agreement with MCFM for the two sub-leading jets at the Tevatron (see Fig. 10.12). The difference in slope with respect to MCFM for the leading p_T^{jet} spectrum was also present at the Tevatron. For ALPGEN+PYTHIA, agreement with MCFM to within 10% is achieved for all three p_T^{jet} spectra by scaling the nominal μ_R by 0.2. As a comparison, for Tevatron a scaling the nominal μ_R by 0.5 (0.3) was found to give good agreement with MCFM (data) (see Fig. 10.11).

11.2 Using $\Delta\phi(\text{dielectron, jet})$ to constrain the Cross Section for Multiple Parton Interactions

The large differences between the predictions from SHERPA and ALPGEN+PYTHIA tune QW of the amount of jets with $p_T < 50$ GeV arising in Z/γ^* events due to multiple parton interactions indicates that early measurements which can help to constrain the amount of jets arising due to such interactions are needed. Several planned searches for new particles at the LHC are sensitive to the production rates of such low- p_T jets, for instance searches for the Higgs boson in the so-called vector boson fusion channel where the reconstruction of jets down to 20 GeV is foreseen [8]. One possible observable which can be used to constrain the cross section for jet production from multiple parton interactions is $\Delta\phi(\text{dielectron, jet})$, inclusive in jets. Jets from multiple parton interactions are expected to have only a small correlation with the dielectron system, and such jets should therefore give a flat contribution

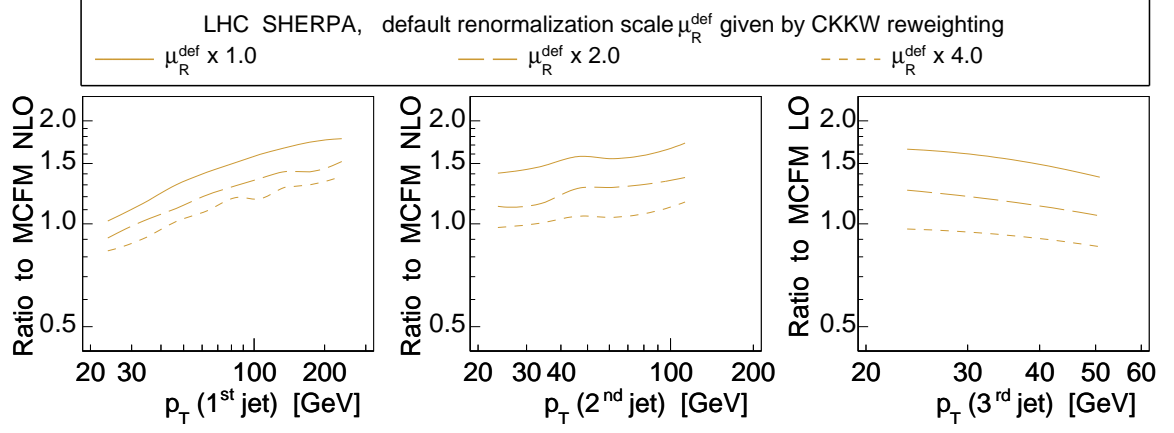


Figure 11.4: Ratios of the LHC predictions from SHERPA to those from MCFM, corresponding to the nominal choice of renormalization scale in SHERPA, as well as to the nominal choice multiplied with a factor of 2 or 4.

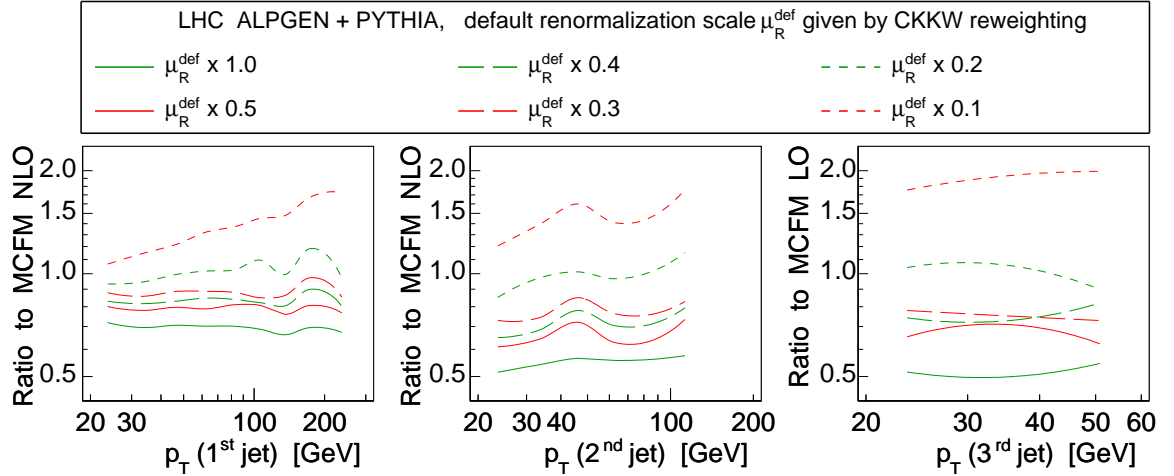


Figure 11.5: Ratios of the LHC predictions from ALPGEN+PYTHIA to those from MCFM, corresponding to the nominal choice of renormalization scale in ALPGEN+PYTHIA, as well as to the nominal choice multiplied with a factor of 0.5, 0.4, 0.3, 0.2 or 0.1.

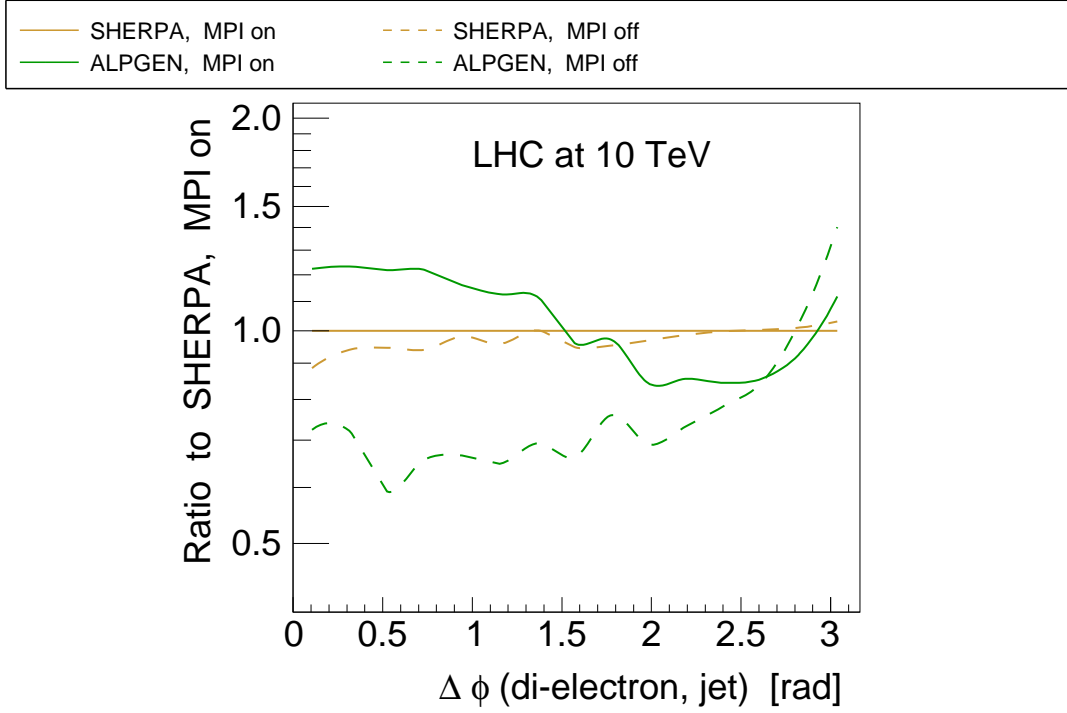


Figure 11.6: The normalized distribution of the jet-inclusive $\Delta\phi(\text{dielectron, jet})$ observable at LHC, as predicted by SHERPA and ALPGEN+PYTHIA both with and without the simulation of multiple parton interactions.

to $\Delta\phi(\text{dielectron, jet})$. The shape of this distribution is therefore sensitive to the amount of jets arising from multiple parton interactions. The SHERPA and ALPGEN+PYTHIA predictions for the inclusive $\Delta\phi(\text{dielectron, jet})$ distribution at LHC, with integral normalized to unity, are shown in Fig. 11.6 with and without the simulation of multiple parton interactions. The event selection used is the same as in the measurement presented in Sect. 9.22. For SHERPA, turning off the multiple parton interaction model changes the observable by less than 10% over the whole range $[0, \pi]$. In contrast, the ALPGEN+PYTHIA prediction decreases by a factor of about two for $\Delta\phi < 1.5$ if the multiple parton interaction model is deactivated. As was seen in the measurement presented in Chapter 9, the shape of the normalized, jet-inclusive $\Delta\phi(\text{dielectron, jet})$ observable has a significantly lower sensitivity to the uncertainty of the jet energy-scale calibration than p_T^{jet} measurements. The calibration of jet energies is expected to be the main source of systematic uncertainty to jet measurements in the ATLAS experiment at the LHC [8], in particular in the start-up phase. Assuming that the perturbative contribution to this observable can be sufficiently well controlled, $\Delta\phi(\text{dielectron, jet})$ is a promising observable for constraining the cross section for the associated production of Z/γ^* and jets arising from multiple parton interaction at an early stage of the LHC experiments.

Chapter 12

Summary and Conclusions

In this thesis several new measurements of the properties of jets produced in association with a Z/γ^* boson in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV have been presented. The analyzed data sample has an integrated luminosity of 1.04 fb^{-1} and was registered by the DØ experiment located at the Tevatron collider. The cross section for $Z/\gamma^*(\rightarrow e^+e^-)+N$ jet production was measured, differential in the transverse momentum of the N^{th} jet in the event ($N \leq 3$), normalized to the inclusive cross section. These measurements represent an improvement over earlier results in terms of reduced experimental uncertainties. The measurement for $N = 3$ is the first of this observable in a hadron collider experiment. In addition, the cross section for $Z/\gamma^*(\rightarrow e^+e^-)+N$ jets ($N \geq 1$) was measured, differential in the difference in azimuthal angle between the di-electron system and any jet in the event and normalized to unity. This is the first measurement of this observable performed at a hadron collider. Since the measurements are fully corrected for the impact of the detector they can be compared directly with theory predictions.

As a fundamental test of QCD the measured p_T^{jet} spectra were compared with the predictions of perturbative calculations at the next-to-leading order in the strong coupling constant, available through the MCFM program. Good agreement between the next-to-leading order predictions and the measurements was observed. This shows that the measurements are in agreement with the Standard Model.

The presented measurements were also used to test the predictions of commonly used event-generator models. SHERPA and ALPGEN+PYTHIA were found to offer a significantly better description of the measurements than MC@NLO, PYTHIA or HERWIG, thereby offering the most accurate particle-level predictions of $Z/\gamma^*(\rightarrow e^+e^-)+\text{jet}$ production currently available. Whereas ALPGEN+PYTHIA was found to give a slightly better description of the shape of the p_T spectrum of the leading jet than SHERPA, the opposite was the case for the shape of the $\Delta\phi(\text{di-electron, jet})$ distribution. Neither generator correctly predicts the overall jet rates. All the studied event generators show a significantly higher sensitivity to the choice of factorization and renormalization scale than the next-to-leading-order calculations. This reflects the leading-order nature of the event generators and corresponds to a reduced predictive power compared with the next-to-leading-order calculations. It was illustrated how the overall jet rates predicted by SHERPA and ALPGEN+PYTHIA can be adjusted a posteriori to improve the agreement with the measurements by multiplying the nominal choice of renormalization scale with a constant factor.

Comparisons between the distributions of the difference in azimuthal angle between the two leading jets in $Z/\gamma^*(\rightarrow e^+e^-)+\text{jets}$ events reconstructed in data and in an event sample generated using the Q^2 -ordered PYTHIA model indicated the presence of a significant excess of events with $\Delta\phi(\text{jet, jet}) = \pi$ in the PYTHIA sample. This excess was found to stem from the model used by Q^2 -ordered PYTHIA to simulate multiple parton interactions, presumably arising from the lack of higher-order corrections to the generated $2 \rightarrow 2$ interactions in the

form of a parton-shower algorithm. SHERPA, using a multiple parton interaction model which includes a parton-shower algorithm, is found to be in good agreement with data. Based on particle-level studies, the same conclusion appears to be valid for the models implemented in JIMMY and in p_T -ordered PYTHIA.

At the LHC the cross sections for the associated production of jets with a Z/γ^* boson will be significantly larger than at the Tevatron. Consequently, it will be important to perform measurements like those presented in this thesis at an early stage. Given the good agreement observed between next-to-leading order calculations and the measurements at Tevatron, it can be assumed that next-to-leading order calculations offer the most reliable predictions of the perturbative part of the p_T^{jet} spectra in $Z/\gamma^* + \text{jet}$ production at the LHC currently available. Motivated by this, it was illustrated how the renormalization scale choice in ALPGEN+PYTHIA and SHERPA can be modified in order to adjust the overall jet rates of these generators to agree with the predictions the next-to-leading-order calculations.

Determining the production rates of jets arising from multiple parton interactions is an important task during the early stages of the LHC experiments. It was demonstrated that a measurement of the shape of the $\Delta\phi(\text{di-electron, jet})$ distribution has the advantage of both high sensitivity to the cross section of this jet production mode, and low sensitivity to the uncertainty of the jet energy scale calibration. Assuming that the perturbative contribution to the observable is sufficiently well controlled, this would allow for an early measurement of the cross section for jet production from multiple parton interactions.

List of Tables

2.1	Overview of the Standard Model fermions and their masses	4
2.2	Fermions of the Standard Model	4
3.1	Simplified summary of the various models for simulating Z/γ^* production . .	35
6.1	Information on the simulated event samples used in chapters 7 and 9.	62
7.1	Number of events for different jet multiplicities measured in data compared with the predictions of PYTHIA and SHERPA.	68
9.1	Background prediction for the jet-inclusive sample	89
9.2	Background prediction for $p_T(\text{leading jet})$	89
9.3	Background prediction for $p_T(\text{second jet})$	89
9.4	Background prediction for $p_T(\text{third jet})$	89
9.5	Cross-check of inclusive cross section	102
9.6	Uncertainties of the measurement of $p_T(\text{leading jet})$	131
9.7	Uncertainties of the measurement of $p_T(\text{second jet})$	131
9.8	Uncertainties of the measurement of $p_T(\text{third jet})$	131
9.9	Summary of the $p_T(\text{leading jet})$ measurement	132
9.10	Summary of the $p_T(\text{second jet})$ measurement	133
9.11	Summary of the $p_T(\text{third jet})$ measurement	133
9.12	Comparison with result from Ref. [128]	133
9.13	Measurement of $\sigma^{-1}d\sigma/d\Delta\phi(\text{di-electron,jet})$	144
10.1	MPI and hadronization corrections for $p_T(\text{leading jet})$	147
10.2	MPI and hadronization corrections for $p_T(\text{second jet})$	147
10.3	MPI and hadronization corrections for $p_T(\text{third jet})$	148

List of Figures

2.1	Standard Model Higgs Potential	8
2.2	Leading-order fermion-fermion scattering diagrams	11
2.3	Betafunction Loop-Diagrams	11
2.4	Running of α_s : pQCD compared with measurements	13
2.5	Tevatron limits on the Higgs production cross section	14
2.6	$\Delta\chi^2$ from a fit to electroweak precision measurement	15
2.7	Standard Model Higgs Branching Ratios	16
2.8	Running coupling constants in the SM and in a MSSM scenario	17
3.1	Drell-Yann Feynman diagrams: LO, NLO real and virtual	19
3.2	p_T^Z : NLO and NNLO vs Data	21
3.3	p_T^Z : NLO+NLL vs Data	24
3.4	Jets at the parton, particle and detector level	31
4.1	The FNAL accelerator chain	38
4.2	Generic Multipurpose Detector	39
4.3	The DØ Run II detector	40
4.4	The central tracking system	41
4.5	Geometry of the SMT system	42
4.6	Calorimeter projective tower structure	43
4.7	Schematic overview of the DØ trigger and data acquisition system	44
5.1	Distribution of four electron likelihood input variables	51
5.2	Distribution of the electron likelihood discriminant	51
5.3	Electron identification efficiencies	52
5.4	Jet identification efficiencies	55
5.5	Jet Energy Correction Correction Factors	57
5.6	Total Jet Energy Scale Correction for Data	58
5.7	ΔS distributions for γ +jet events	59
5.8	Mean and width of the ΔS distributions for Z/γ^* +jet events	59
7.1	Electron identification efficiency correcion	64
7.2	Jet energy resolution is Data and Simulation	65
7.3	Di-electron invariant mass distributions	66
7.4	Di-electron mass distributions for fiducial and non-fiducial electrons	66
7.5	Di-electron mass distributions with and without QED final-state radiation	67
7.6	SHERPA and PYTHIA vs data: di-electron p_T	68
7.7	SHERPA and PYTHIA vs data: jet multiplicity	69
7.8	SHERPA and PYTHIA vs data: p_T leading jet	70
7.9	SHERPA and PYTHIA vs data: p_T second jet	71
7.10	SHERPA and PYTHIA vs data: p_T third jet	72

7.11	SHERPA and PYTHIA vs data: $\Delta\eta(\text{jet}, \text{jet})$	73
7.12	SHERPA and PYTHIA vs data: $\Delta\phi(\text{jet}, \text{jet})$	74
7.13	SHERPA and PYTHIA vs data: p_T third jet with VBF-like selection	75
7.14	SHERPA and PYTHIA vs data: relative η of the third jet with VBF-like selection	76
7.15	SHERPA p_T^{jet} spectra: v1.0.6 vs v1.0.8	77
8.1	Migration Matrix Inversion Problem	83
8.2	Bin-by-bin unfolding versus migration-matrix unfolding	85
9.1	Same-sign sample M(e,e) distribution	90
9.2	Same-sign sample M(e,e) distribution vs p_T^{jet}	91
9.3	Photons from Z decays	93
9.4	Electron p_T at particle vs detector level	94
9.5	Particle-level jet radial momentum density	95
9.6	Particle-level jet composition	96
9.7	Correlation between p_T^{jet} at particle level and detector level	97
9.8	Correlation between p_T^{jet} at particle level and detector level	98
9.9	$p_T(\text{leading jet})$, data vs ALPGEN+PYTHIA	98
9.10	Instantaneous luminosity reweighting	99
9.11	Vertex z position reweighting	100
9.12	EM energy scale correction	101
9.13	ΔS for $25 < p_T^Z < 30$ GeV and $60 < p_T^Z < 70$ GeV	105
9.14	Mean and RMS of ΔS	105
9.15	Mean and RMS of ΔS	106
9.16	Mean and RMS of ΔS for $N_{\text{jet}} \geq 1$	106
9.17	Mean and RMS of ΔS for $N_{\text{jet}} \geq 2$	107
9.18	ΔS for $N_{\text{jet}} \geq 3$	107
9.19	Electron trigger efficiency in jet events	108
9.20	Electron trigger efficiency in jet events vs p_T^e	109
9.21	Minimal ΔR between electrons and jets	111
9.22	Scalar p_T sum of tracks in N -jet and 0-jet events	113
9.23	Average p_T of tracks in N -jet and 0-jet events	113
9.24	Simulation vs data before kinematic reweighting	114
9.25	Data vs ALPGEN+PYTHIA, after p_T^Z reweighting	116
9.26	p_T^Z vs p_T^{jet} : ALPGEN+PYTHIA vs SHERPA	117
9.27	p_T^Z vs $\Delta\phi(Z, \text{leading jet})$: ALPGEN+PYTHIA vs SHERPA	118
9.28	Simulation vs data after 2D re-weighting to SHERPA	119
9.29	Simulation vs data after 2D re-weighting to SHERPA	121
9.30	Low- p_T bias from p_T^{jet} smearing	122
9.31	p_T^{jet} reweighting	123
9.32	η^{jet} reweighting	124
9.33	Impact of $(N-1)^{\text{th}}$ and $(N+1)^{\text{th}}$ jets on $p_T(N^{\text{th}} \text{ jet})$	125
9.34	$(\mathcal{A} \times \epsilon)_{\text{sim}}$ for $p_T(\text{three leading jets})$	125
9.35	Efficiency of $\Delta R(e, \text{jet}) > 0.4$ selection	126
9.36	Efficiency of lepton selection	127
9.37	Uncertainties $p_T(\text{leading jet})$	128
9.38	Uncertainties $p_T(\text{second jet})$	129

9.39	Uncertainties p_T (third jet)	130
9.40	Comparison with result from Ref. [129]	134
9.41	Comparisons of uncertainties with Ref. [130]	135
9.42	Studies on $\Delta\phi$ (di-electron, leading jet)	137
9.43	Di-electron direction of flight	138
9.44	Studies on $\Delta\phi$ (di-electron, all jet)	139
9.45	Particle-level jet confirmation	140
9.46	Anti pile-up jet cuts	141
9.47	Particle-level jet confirmation after anti pile-up selection	141
9.48	Studies on jet-inclusive $\Delta\phi$ (di-electron, jet), after pile-up removal	142
9.49	$\Delta\phi$, data vs ALPGEN+PYTHIA	143
9.50	$(\mathcal{A} \times \epsilon)$ for $\Delta\phi$ (di-electron, jet)	144
10.1	p_T^{jet} spectra: data vs MCFM	149
10.2	p_T^{jet} spectra: NLO MCFM vs data	149
10.3	p_T^{jet} spectra: LO MCFM vs data	150
10.4	p_T^{jet} spectra: SHERPA vs data	151
10.5	p_T^{jet} spectra: ALPGEN+PYTHIA vs data	151
10.6	p_T^{jet} spectra: MC@NLO vs data	152
10.7	p_T^{jet} spectra: Q^2 -ordered PYTHIA vs data	153
10.8	p_T^{jet} spectra: p_T -ordered PYTHIA vs data	153
10.9	p_T^{jet} spectra: HERWIG+JIMMY vs data	153
10.10	p_T^{jet} spectra: comparisons between the various event generators	154
10.11	p_T^{jet} spectra: tuning μ_R -choice for ALPGEN+PYTHIA	155
10.12	p_T^{jet} spectra: tuning μ_R -choice for SHERPA	156
10.13	p_T^{jet} spectra: the impact of the D parameter in SHERPA	157
10.14	p_T^{jet} spectra: the impact of the p_T^{cut} parameter in SHERPA	157
10.15	$\Delta\phi$ (di-electron,jet): SHERPA vs data	158
10.16	$\Delta\phi$ (di-electron,jet): data compared with various event generator models	159
10.17	$\Delta\phi$ (di-electron,jet): ALPGEN+PYTHIA split into parton bins	159
10.18	$\Delta\phi$ (jet,jet): event generator predictions with and without MPI	160
11.1	p_T^{jet} spectra at LHC vs at Tevatron: MCFM	162
11.2	p_T^{jet} spectra at LHC vs at Tevatron: SHERPA	163
11.3	p_T^{jet} spectra at LHC vs at Tevatron: SHERPA	164
11.4	SHERPA p_T^{jet} predictions for various choices for μ_R	165
11.5	ALPGEN+PYTHIA p_T^{jet} predictions for various choices for μ_R	165
11.6	$\Delta\phi$ (dielectron,jet) at LHC	166

Bibliography

- [1] S. Eidelman et al. Review of Particle Physics. *Phys. Lett.*, B592:1, 2004.
- [2] Y. Fukuda et al. Evidence for Oscillation of Atmospheric Neutrinos. *Phys. Rev. Lett.*, 81:1562, 1998, hep-ex/9807003.
- [3] Rajan Gupta. Introduction to lattice QCD. 1997, hep-lat/9807028.
- [4] Siegfried Bethke. Experimental Tests of Asymptotic Freedom. *Prog. Part. Nucl. Phys.*, 58:351–386, 2007, hep-ex/0606035.
- [5] G. Abbiendi. Search for the Standard Model Higgs Boson at LEP. *Phys. Lett.*, B565:61, 2003.
- [6] Tevatron New Phenomena Higgs Working Group. Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with up to 4.2 fb^{-1} of Data. 2009, 0903.4001.
- [7] Gfitter group. <https://project-gfitter.web.cern.ch/project-gfitter/GSM/>.
- [8] G. Aad et al. Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics. 2009, 0901.0512.
- [9] J. Butterworth, A. Davidson, M. Rubin, and G. Salam. Jet substructure as a new Higgs search channel at LHC. *Phys. Rev. Lett.*, 100:242001, 2008, 0802.2470.
- [10] A. Djouadi, J. Kalinowski, and M. Spira. HDECAY: A program for Higgs boson decays in the standard model and its supersymmetric extension. *Comput. Phys. Commun.*, 108:56–74, 1998, hep-ph/9704448.
- [11] Maxim Pospelov and Adam Ritz. Electric dipole moments as probes of new physics. *Annals Phys.*, 318:119–169, 2005, hep-ph/0504231.
- [12] E. Komatsu et al. Five-Year Wilkinson Microwave Anisotropy Probe: WMAP Observations and Cosmological Interpretation. *Astrophys. J. Suppl.*, 180:330–376, 2009, 0803.0547.
- [13] S.P. Martin. A Supersymmetry Primer, 2006. hep-ph/9709356v4.
- [14] Nima Arkani-Hamed, Andrew G. Cohen, and Howard Georgi. Electroweak symmetry breaking from dimensional deconstruction. *Phys. Lett.*, B513:232–240, 2001, hep-ph/0105239.
- [15] Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. The hierarchy problem and new dimensions at a millimeter. *Phys. Lett.*, B429:263–272, 1998, hep-ph/9803315.

- [16] R. D. Peccei and Helen R. Quinn. CP Conservation in the Presence of Instantons. *Phys. Rev. Lett.*, 38:1440–1443, 1977.
- [17] CERN Axion Solar Telescope. <http://cast.web.cern.ch/CAST/index.html>.
- [18] T. Kinoshita. Mass singularities of Feynman amplitudes. *J. Math. Phys.*, 3:650–677, 1962.
- [19] Kirill Melnikov and Frank Petriello. Electroweak gauge boson production at hadron colliders through $O(\alpha(s)^2)$. *Phys. Rev.*, D74:114017, 2006.
- [20] John M. Campbell and R. Keith Ellis. Next-to-leading order corrections to $W + 2\text{jet}$ and $Z + 2\text{jet}$ production at hadron colliders. *Phys. Rev.*, D65:113007, 2002.
- [21] R. Keith Ellis, Kirill Melnikov, and Giulia Zanderighi. Generalized unitarity at work: first NLO QCD results for hadronic $W + 3\text{jet}$ production. 2009, 0901.4101.
- [22] R. Keith Ellis, Kirill Melnikov, and Giulia Zanderighi. $W+3$ jet production at the Tevatron. 2009, 0906.1445.
- [23] C. F. Berger et al. Precise Predictions for $W + 3$ Jet Production at Hadron Colliders. 2009, 0902.2760.
- [24] M. L. Mangano et al. ALPGEN, a Generator for Hard Multiparton Processes in Hadronic Collisions. *JHEP*, 0307:001, 2003, hep-ph/0206293.
- [25] Tanju Gleisberg and Stefan Hoche. Comix, a new matrix element generator. *JHEP*, 12:039, 2008, 0808.3674.
- [26] G. Bozzi et al. Transverse-momentum resummation: a perturbative study of Z production at the Tevatron. 2008, hep-ph/0812.2862.
- [27] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys.*, B126:298, 1977.
- [28] V. V. Sudakov. Vertex parts at very high-energies in quantum electrodynamics. *Sov. Phys. JETP*, 3:65–71, 1956.
- [29] D. Amati, A. Bassetto, M. Ciafaloni, G. Marchesini, and G. Veneziano. A Treatment of Hard Processes Sensitive to the Infrared Structure of QCD. *Nucl. Phys.*, B173:429, 1980.
- [30] Torbjorn Sjostrand. A Model for Initial State Parton Showers. *Phys. Lett.*, B157:321, 1985.
- [31] M. Bengtsson and T. Sjostrand. A Comparative Study of Coherent and Noncoherent Parton Shower Evolution. *Nucl. Phys. B*, 289:810, 1987.
- [32] M.H. Seymour. Matrix element corrections to parton shower algorithms. *Comput. Phys. Comm.*, 90:95, 1995.
- [33] M.H. Seymour. A Simple prescription for first order corrections to quark scattering and annihilation processes. *Nucl. Phys. B*, 436:443, 1995.

-
- [34] Stefano Frixione and Bryan R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP*, 06:029, 2002, hep-ph/0204244.
 - [35] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. The POWHEG-hvq manual version 1.0. 2007, 0707.3081.
 - [36] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber. QCD Matrix Elements + Parton Showers. *JHEP*, 11:063, 2001, hep-ph/0109231.
 - [37] F. Krauss. Matrix elements and parton showers in hadronic interactions. *JHEP*, 08:015, 2002, hep-ph/0205283.
 - [38] T. Gleisberg et al. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009, 0811.4622.
 - [39] Gerald C. Blazey et al. Run II jet physics. 2000, hep-ex/0005012.
 - [40] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, and Michele Treccani. Matching matrix elements and shower evolution for top- quark production in hadronic collisions. *JHEP*, 01:013, 2007, hep-ph/0611129.
 - [41] J. Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J.*, C53:473–500, 2008, 0706.2569.
 - [42] W. Bartel et al. Test of Fragmentation Models by Comparison with Three Jet Events Produced in $e^+ e^- \rightarrow \text{Hadrons}$. *Phys. Lett.*, B134:275, 1984.
 - [43] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys. Rept.*, 97:31–145, 1983.
 - [44] B. R. Webber. A QCD Model for Jet Fragmentation Including Soft Gluon Interference. *Nucl. Phys.*, B238:492, 1984.
 - [45] Pavel M. Nadolsky et al. Implications of CTEQ global analysis for collider observables. *Phys. Rev.*, D78:013004, 2008, 0802.0007.
 - [46] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Uncertainties on α_s in global PDF analyses. 2009, 0905.3531.
 - [47] V. N. Gribov and L. N. Lipatov. Deep inelastic ep scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:438–450, 1972.
 - [48] Yuri L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and $e^+ e^-$ Annihilation by Perturbation Theory in Quantum Chromodynamics. (In Russian). *Sov. Phys. JETP*, 46:641–653, 1977.
 - [49] J. Pumplin et al. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 07:012, 2002, hep-ph/0201195.
 - [50] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006, hep-ph/0603175.

- [51] Michael G. Albrow et al. Tevatron-for-LHC Report of the QCD Working Group. 2006, hep-ph/0610012.
- [52] CDF Collaboration. Measurement of double parton scattering in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ tev. *Phys. Rev. Lett.*, 79(4):584, 1997.
- [53] Torbjorn Sjostrand and Maria van Zijl. A Multiple Interaction Model for the Event Structure in Hadron Collisions. *Phys. Rev.*, D36:2019, 1987.
- [54] CDF Collaboration. Underlying Event in Inclusive Jet Production, 2003. <http://www-cdf.fnal.gov/physics/new/qcd/run2/ue/chgjet/index.html>.
- [55] John M. Campbell and R. Keith Ellis. Next-to-leading order corrections to $W+2$ jet and $Z+2$ jet production at hadron colliders. *Phys. Rev.*, D65:113007, 2002, hep-ph/0202176.
- [56] G. Corcella et al. HERWIG 6.5 release note. 2002, hep-ph/0210213.
- [57] J. M. Butterworth, Jeffrey R. Forshaw, and M. H. Seymour. Multiparton interactions in photoproduction at HERA. *Z. Phys.*, C72:637–646, 1996, hep-ph/9601371.
- [58] M. Voutilainen. Measurements of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV, 2008. PhD thesis.
- [59] Gavin P. Salam and Gregory Soyez. A Practical Seedless Infrared-Safe Cone jet algorithm. *JHEP*, 05:086, 2007, 0704.0292.
- [60] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev.*, D48:3160–3166, 1993, hep-ph/9305266.
- [61] S. Catani, Yuri L. Dokshitzer, M. H. Seymour, and B. R. Webber. Longitudinally invariant K_t clustering algorithms for hadron hadron collisions. *Nucl. Phys.*, B406:187–224, 1993.
- [62] Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. Better Jet Clustering Algorithms. *JHEP*, 08:001, 1997, hep-ph/9707323.
- [63] M. Wobisch and T. Wengler. Hadronization corrections to jet cross sections in deep-inelastic scattering. 1998, hep-ph/9907280.
- [64] Fermilab Beams Division. Run II Handbook, 2003. <http://www-bd.fnal.gov/runII/index.html>.
- [65] Fermi National Accelerator Laboratory. <http://www.fnal.gov>.
- [66] S. Abachi et al. Observation of the Top Quark. *Phys. Rev. Lett.*, 74:2632, 1995, hep-ex/9503003.
- [67] F. Abe et al. Observation of Top Quark Production in $p\bar{p}$ Collisions. *Phys. Rev. Lett.*, 74:2626, 1995, hep-ex/9503002.
- [68] D. Mohl, G. Petrucci, L. Thorndahl, and Simon van der Meer. Physics and Technique of Stochastic Cooling. *Phys. Rep.*, 58:73, 1980.

-
- [69] Fermilab Electron Cooling Project. <http://www-ecool.fnal.gov/>.
 - [70] Tejinder Virdee. Inaugural Lecture.
 - [71] DØ Collaboration. The Upgraded DØ Detector. *Nucl. Instrum. Meth.*, A565:463, 2006.
 - [72] R. P. Smith. Describing the DØ solenoidal in DØgeant. DØ Note 4803.
 - [73] DØ Collaboration. <http://www-d0.fnal.gov/D0Code/source/d0reco>.
 - [74] R. Brun and F. Rademakers. ROOT: An Object Oriented Data Analysis Framework. *Nucl. Instrum. Meth.*, A389:81, 1997.
 - [75] DØ Data Format Working Group. Recommendations Regarding Common Analysis Format Content. DØ Note 4647.
 - [76] DØ Common Samples Group. <http://www-d0.fnal.gov/Run2Physics/cs/index.html>.
 - [77] CERN Computing Application Software Group and Networks Division. GEANT, Detector Description and Simulation Tool, 2003. CERN Program Library Long Writeup W5013.
 - [78] Y. Fisyak and J. Womersley. DØGSTAR, DØ GEANT Simulation of the Total Apparatus Response. DØ Note 3191.
 - [79] DØ Collaboration. <http://www-d0.fnal.gov/computing/MonteCarlo/simulation/d0sim.html>.
 - [80] H. Greenlee. The DØ Kalman Track Fit. DØ Note 4303.
 - [81] A. Schwartzman, C. Tully. Primary vertex reconstruction by means of adaptive vertex fitting. CERN CMS-NOTE-2007-008.
 - [82] R. Fuhwirth, W. Waltenberger, P. Vanlaer. Adaptive Vertex Fitting. CERN CMS-NOTE-2007-008.
 - [83] G. Bernandi J.-R. Vlimant, U. Bassler and S. Trincz-Duviod. Technical description of the T42 Algorithm for the calorimeter noise suppression. DØ Note 4146.
 - [84] E. Busato G. Bernandi and J.-R. Vlimant. Improvements from the T42 Algorithm on Calorimeter Objects Reconstruction. DØ Note 4335.
 - [85] J. Mitrevski C. Schwanenberger L. Wang, J. Hays. Electron likelihood efficiency in p17. DØ Note 5114.
 - [86] U. Blumenschein. Search for the Associated Production of Charginos and Neutralinos in $p\bar{p}$ Collisions at 1.96 TeV, 2004. PhD thesis.
 - [87] C. Jarvis J. Mans T. Toole M. Yan A. Baden, S. Eno. Measurement of the shape of the boson rapidity distribution for $p\bar{p} \rightarrow Z/\gamma^* \rightarrow e^+e^- + X$ events produced at \sqrt{s} of 1.96 TeV. DØ Note 5247.
 - [88] EM-ID Group. <https://plone4.fnal.gov/P1/D0Wiki/object-id/emid/>.

- [89] C. Schwanenberger T. Toole J. Hayes, J. Mitrevski. Single Electron Efficiencies in p17 Data and Monte-Carlo. DØ Note 5025.
- [90] I. M. Adam. Measurement of the W Boson Mass with the DØ Detector using the Electron E_T Spectrum, 1997. PhD Thesis.
- [91] Q. Zhu. Measurement of the W Boson Mass in Proton-Antiproton Collisions at $\sqrt{s} = 1.8$ TeV, 1994. PhD Thesis.
- [92] M.-C. Cousinou A. Cothnet S. Kermiche, A. Mendes. Energy Scale studies and calibration of the DØ electromagnetic calorimeter using Z^0 and $J/\Psi \rightarrow e^+e^-$ Run II events. DØ Note 4945.
- [93] J. Zhu. Determination of the Electron Energy Scale and Energy Resolution using P14 $Z \rightarrow e^+e^-$ data. DØ Note 4323.
- [94] A. Melntchouk T. Andeen. Measurement of Electron Energy Scale and Offset ith $Z \rightarrow ee$ Full Monte Carlo and Data. DØ Note 5662.
- [95] A. Harel. Jet ID Optimization. DØ Note 4919.
- [96] H. Nogima M. Rangel M. Voutilainen B. Andrieu, A. Harel. Measuring Reconstruction*Jet-ID efficiencies using the tag and probe method in p17. DØ Note 5250.
- [97] M. Rangel M. Voutilainen A. Harel, H. Nogima. Combined JetID efficiency. DØ Note 5218.
- [98] DØ Jet Energy Scale Group. Jet Energy Scale Determination at DØ Run II (Final p17 Version). DØ Note 5382.
- [99] C. Ochando and J.-F. Grivaz. SSR for p17. DØ Note 5609.
- [100] N. Makovec and J.-F. Grivaz. Shifting, Smearing and Removing Simulated Jets. DØ Note 4914.
- [101] DØ Common Sample Group. <http://www-d0.fnal.gov/Run2Physics/cs/index.html>.
- [102] T. Edwards et al. Determination of the Effective Inelastic $p\bar{p}$ Cross Section for the DØ Run II Luminosity Measurement. FERMILAB-TM-2278-E.
- [103] DØ Jet Energy Scale Group. Preliminary jes, v07-01-01.
- [104] S. Schumann. Private communication.
- [105] C. Buttar et al. Standard Model Handles and Candles Working Group: Tools and Jets Summary Report. 2008, 0803.0678.
- [106] V. M. Abazov et al. Measurement of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. Lett.*, 101:062001, 2008, 0802.2400.
- [107] Andreas Hocker and Vakhtang Kartvelishvili. SVD Approach to Data Unfolding. *Nucl. Instrum. Meth.*, A372:469–481, 1996, hep-ph/9509307.

-
- [108] Volker Blobel. An unfolding method for high energy physics experiments. 2002, hep-ex/0208022.
- [109] A. Abulencia et al. Measurements of Inclusive W and Z Cross Sections in p-pbar Collisions at $\sqrt{s}=1.96$ TeV. *J. Phys.*, G34:2457–2544, 2007, hep-ex/0508029.
- [110] A. Abulencia et al. 2006. <http://www-cdf.fnal.gov/physics/ewk/2006/zmm/index.html>.
- [111] R. Hamberg, W. L. van Neerven, and T. Matsuura. A Complete calculation of the order $\alpha - s^2$ correction to the Drell-Yan K factor. *Nucl. Phys.*, B359:343–405, 1991.
- [112] R. Hamberg, W. L. van Neerven, and T. Matsuura. A Complete calculation of the order α_s^2 correction to the Drell-Yan K factor. *Nucl. Phys.*, B359:343, 1991. [Erratum-ibid. B644:402, 2002].
- [113] T. Nunnemann. NNLO Cross Sections for Drell-Yan, Z and W Production using Modern Parton Distribution Functions. DØ Note 4476.
- [114] H. Schellman. The longitudinal shape of the luminous region at DØ. DØ Note 5142.
- [115] V. M. Abazov et al. Measurement of the shape of the boson rapidity distribution for $p\bar{p} \rightarrow Z/\gamma^* \rightarrow e^+e^- + X$ events produced at \sqrt{s} of 1.96-TeV. *Phys. Rev.*, D76:012003, 2007, hep-ex/0702025.
- [116] S. Park, M. Biegel. Efficiency of the Data Quality Calorimeter Flags. DØ Note 5324.
- [117] H. Fox and others. Measurement of the cross section for $Z/\gamma^* \rightarrow e^+e^-$ production at DØ. DØ Note 5627.
- [118] V. M. Abazov et al. Measurement of the shape of the boson transverse momentum distribution in $p\bar{p} \rightarrow Z/\gamma^* \rightarrow e^+e^- + X$ events produced at $\sqrt{s} = 1.96$ -TeV. *Phys. Rev. Lett.*, 100:102002, 2008, 0712.0803.
- [119] DØ Collaboration. Z + jet production in the DØ experiment: A comparison between data and the PYTHIA and SHERPA Monte Carlos. DØ Note 5066-CONF.
- [120] S. Alekhin et al. HERA and the LHC - A workshop on the implications of HERA for LHC physics: Proceedings Part A. 2005, hep-ph/0601012.
- [121] Anthony Allen Affolder et al. Charged jet evolution and the underlying event in $p\bar{p}$ collisions at 1.8 TeV. *Phys. Rev.*, D65:092002, 2002.
- [122] W. Fisher, J. Haley, J. Sekaric. A measurement of Z boson+Jets differential production cross sections‘A Measurment of Di-boson Production in Lepton Plus Jet Decays. DØ Note 5544.
- [123] M. Shamim, T. Bolton. Generator Level Reweighting of pT of Z Boson. DØ Note 5565.
- [124] J. Pumplin et al. Uncertainties of predictions from parton distribution functions. 2. The Hessian method. *Phys. Rev.*, D65:014013, 2001, hep-ph/0101032.

- [125] B. Tiller and T. Nunnemann. A measurement of Z boson+Jets differential production cross sections. DØ Note 5543.
- [126] G. D. Lafferty and T. R. Wyatt. Where to stick your data points: The treatment of measurements within wide bins. *Nucl. Instrum. Meth.*, A355:541–547, 1995.
- [127] D. de Florian and M. Grazzini. Higgs production through gluon fusion: updated cross sections at the tevatron and the lhc. arXiv: hep-ph/0901.2427.
- [128] D0 Collaboration and V. Abazov. Measurement of the ratios of the $z/g^* + \geq n$ jet production cross sections to the total inclusive z/g^* cross section in ppbar collisions at $\sqrt{s} = 1.96$ tev. *Physics Letters B*, 658:112, 2008.
- [129] D0 Collaboration: V. Abazov. Measurement of differential $z/\gamma^* + \text{jet} + x$ cross sections in proton anti-proton collisions at $\sqrt{s} = 1.96$ tev. *Physics Letters B*, 669:278, 2008.
- [130] T. Aaltonen et al. Measurement of inclusive jet cross-sections in $Z/\gamma^*(\rightarrow e^+e^-)$ + jets production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ -TeV. *Phys. Rev. Lett.*, 100:102001, 2008, 0711.3717.
- [131] A. Abulencia et al. 2008. http://www-cdf.fnal.gov/physics/new/qcd/zjets_08/tables.pdf.
- [132] M. R. Whalley, D. Bourilkov, and R. C. Group. The Les Houches Accord PDFs (LHAPDF) and Lhaglu. 2005, hep-ph/0508110.
- [133] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006, hep-ph/0603175.
- [134] T. Sjostrand and P. Z. Skands. Transverse-momentum-ordered showers and interleaved multiple interactions. *Eur. Phys. J.*, C39:129–154, 2005, hep-ph/0408302.
- [135] Peter Skands and Daniel Wicke. Non-perturbative QCD effects and the top mass at the Tevatron. *Eur. Phys. J.*, C52:133–140, 2007, hep-ph/0703081.
- [136] M. Wobisch. Private communication.
- [137] M. H. Seymour. Private communication.
- [138] D. Maitre. Private communication.

Acknowledgements

First of all I would like to thank Karl Jakobs for giving me the opportunity to come to Freiburg and perform my PhD studies as a member of his group. Thank you for always being available with solutions and motivations on short notice whenever I got stuck.

Thank you to Prof. Dr. G. Herten and Prof. Dr. S. Dittmaier for agreeing to take part in the PhD committee, and to the Koreferent, whose identity is unknown to me at the time of writing, for evaluating the thesis.

I'm most grateful to Harald, Volker and Ralf for all the advice, help and many interesting discussions.

Many thanks to all my colleagues in the DØ experiment, in particular to the members of the DØ QCD group and EB-002 for triggering many interesting studies and discussions.

I'm highly indebted to Steffen for his many hours of trying to de-confuse my confused mind – thank you for your patience and our many pleasant conversations. Many thanks to Frank, and his team, for excellent user support and for hosting a highly memorable stay in Dresden. Also, I'd like to thank Alberto and Jürgen, our local theorists, for having time for an experimentalist in need of help.

Olav and Robindra, too bad that our Norwegian Colony was split apart: it was a nice time! Ingo, it is always a pleasure to spend time with you, whether in the office or far, far away from it. Thank you, Giacinto and Michael, for the nice company and interesting discussions, and Uli, for many nice evenings and excursions! Many thanks to the legendary M&M-team and to Giacinto for offering 24/7 computer and programming support! Also thanks to Björn, for all the times when you got freiburg-clued0 up running again. Xavier, Alberto and Christian, I'm grateful for the valuable last-minute help with proof-reading. Frank, Jörg, Inga, Sacha, Urban and Jochen all have one thing in common: having been my office mates during shorter or longer parts of the last four years. We might not always have had the cleanest, tidiest office with the largest density of plants, but I've enjoyed your company a lot. And we've definitely had the nicest Christmas train and the best coffee! Maxim, it was a pleasure to share the experience of Moriond-rush with you. I'd like to thank the entire ATLAS and Medipix parts of the group for interesting discussions, for many nice lunches, and for a pleasant atmosphere in general. And apologies to everyone I've forgot to mention in person – my memory has never been the best, and in three hours Frau Seger expects a visit from me.

Michael and Natalia, thank you for many nice evenings, holidays and visits to the Straussen of Süd-Baden, and for you hospitality in Frankfurt.

Thank you, Chris, for always helping a confused physicist with formalities and paper-work.

Many thanks to the Reader, for getting this far. I assume you have worked your way carefully through all chapters and did not skip immediately to this section.

I am very grateful to my family, Tor, Frøydis, Magnus, Siw, Mormor, Bestefar and Farmor, for all the support and encouragement, and for always being there for me.

Finally, Sonja, I want to thank you for your love and understanding, and for encouraging and supporting me whenever I needed it.